INTERFERENCE AND VOLATILITY IN EVOLUTIONARY AGENT-BASED SYSTEMS

CHLOE MAY BARNES

Doctor of Philosophy

ASTON UNIVERSITY

August 2021

Abstract

Agents that exist and pursue individual goals in shared environments can indirectly affect one another in unanticipated ways, such that the actions of others in the environment can interfere with the ability to achieve goals. Despite this, the impact that these unintended interactions and interference can have on agents is not currently well understood. This is problematic as these goal-oriented agents are increasingly situated in complex sociotechnical systems, that are composed of many actors that are heterogeneous in nature.

The primary aim of this thesis is to explore the effect that indirect interference from others has on evolution and goal-achieving behaviour in agent-based systems. More specifically, this is investigated in the context of agents that do not possess the ability to perceive or learn about others within the environment, as information about others may not be readily available at runtime, or there may be a distinct lack of capacity to obtain such information. By conducting three experimental studies, it is established that evolutionary volatility is a consequence of indirect interactions between goal-oriented agents in a shared environment, and that these consequences can be mitigated by designing more socially-sensitive agents. Specifically, agents that employ social action are demonstrated to reduce the evolutionary volatility experienced by goal-oriented agents, without affecting the fitness received. Additionally, behavioural plasticity achieved via neuromodulation is shown to allow coexisting agents to achieve their goals more often with less evolutionary volatility in highly variable environments. While sufficient approaches to mitigate interference include learning about or modelling others, or for agents to be explicitly designed to identify interference to mitigate its consequences, this thesis demonstrates that these are not necessary. Instead, more socially-sensitive agents are shown to be capable of achieving their goals and mitigating interference without this knowledge of others, simply by shifting the focus from goal-oriented actions to more socially-oriented behaviour.

To all who have inspired and encouraged me,
and to those I hope to inspire in the future:
Thank you for believing in me.

Acknowledgements

Words cannot express the immense gratitude I have for my supervisors, Dr Peter Lewis and Dr Anikó Ekárt, who have been monumental inspirations to me during my studies. My first lecture at Aston was taught by Peter – his first at the university as well; little did I know about what was to come in the next few years! His enthusiasm for research and computational intelligence has been contagious, and I am thankful for both the professional and life advice he has given to me during my time at Aston. One of the things I think I am most grateful for is his guidance 'through the fog' when trying to find my place in the research world – but also for helping me to gain the confidence in myself to find my own way in the end. My thanks go to Anikó for inspiring me to pursue a career of research in the first place. Her talk on evolutionary art during a session for prospective students whilst I was in the first year of my undergraduate degree fascinated me, and sparked my interest in computational intelligence. I am grateful for her meticulous guidance, listening ear, and continuous encouragement while completing my studies; she has taught me how to believe in my own abilities, and I will always be grateful to her for inspiring me from the very beginning of my time at Aston. While finishing this thesis seemed like an impossible endeavour at times, both Peter and Anikó's support, encouragement and guidance has helped me to achieve more than I ever could have hoped for during my PhD studies. I have learnt so much from them both; they have shown me that there are always more questions to ask, and that one of the most important questions to ask is *why?*. I hope that one day I can pass on their wisdom to other young researchers, and encourage them to find the best in themselves – as they have both done for me.

I cannot thank Dr Kai Olav Ellefsen, Dr Kyrre Glette, and Professor Jim Tørresen from the ROBIN research group at the University of Oslo enough – without whom my PhD would have looked very, very different. The thought-provoking discussions we had during my visit

to Norway have helped to shape the rest of my studies, and indeed my thesis as well. I am truly grateful to have had the opportunity to work with them – even more so considering that I first visited Oslo at the beginning of the pandemic in 2020; their expertise has been invaluable, and I hope that we can continue our collaboration – one day hopefully over a coffee at Fuglen once more.

I would also like to thank all the brilliant researchers with whom I have had the pleasure of working with in these past few years. In particular, the SISSY, ACSOS and ALife communities have demonstrated to me the importance of interdisciplinary research, and I am grateful to these communities for welcoming myself and other young researchers with open arms into the discussion. I especially would like to thank Dr Lukas Esterle, Dr John Brown, Dr Kirstie Bellman, Dr Chris Landauer, Dr James Borg, Dr Adam Stanton, and Dr Jeremy Pitt, for the wonderful discussions on rivers, superstition, human behaviour, and life in general; their support and advice has been invaluable – especially in the early stages of my PhD when I was still trying to find my feet and my place in the community. In addition to those who have helped me to get to this point, I am also grateful to Professor Peter McBurney and Dr James Borg for their critiques, questions, and discussions when examining my PhD thesis.

Getting to this point would have been impossible without the support and understanding of my friends, whose friendships I have so much to be grateful for. Studying computer science would not have been as enjoyable as it has been without James Langford and Alex Sanghera; these two kept me grounded during our undergraduate studies, and encouraged me right from the very beginning. I am also thankful to Lt Col David King and Alwyn Burger for making my first conference experience memorable in so many ways; I'm so glad we've stayed friends, and I'm looking forward to having a beer in person once again. As well as Alwyn, I'd like to thank Wiebke Busch for reminding me that there's more to life than research, and for still sticking around – especially when I've been in a hole.

Words cannot express how grateful I am to Dr Jordan Bird, and I am so glad that we have been able to ride the roller-coaster that is reading for a PhD together. Without him, the last few years would have felt very empty. His unwavering passion for AI has been a true inspiration to me, and he has continuously helped me to rediscover my love for research in new ways, time and time again. I could not have done this without him, as he has been there for me in both the highs and lows of life and science – for that, I am forever grateful.

It's overwhelming to think of all the things that we've done and achieved over the past few years, and I'm so excited to see what comes next for us both.

My family have supported and encouraged me throughout my life, and have given me the confidence to follow my dreams. I am grateful to my Mum and Dad, Sharon and Phil, who have always taught me to strive to be the best version of myself. They helped me to believe in myself and realise that I can achieve anything that I set my mind to; not only this, but they have always tried to give myself and my sister the best opportunities in life, and for that I am immeasurably grateful. I am so proud of my little sister, Hannah, and all she has achieved, despite the difficulties she has been faced with; she has taught me that dedication, perseverance, and resilience are key to success. I am inspired by her strength and determination, and her motivation to achieve the best she can has helped me to do the same. I have always been taught to *just try my best*, and I couldn't have achieved as much as I have done in my life without my family cheering me on. From the bottom of my heart, thank you.

List of Publications

Publications Arising from this Thesis:

[15] C. M. Barnes, A. Ekárt, and P. R. Lewis. Social Action in Socially Situated Agents. In *Proceedings of the IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 97–106. IEEE, 2019. doi: https://doi.org/10.1109/SASO.2019.00021.

[17] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Coevolutionary Learning of Neuromodulated Controllers for Multi-Stage and Gamified Tasks. In *Proceedings of the IEEE 1st International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 129–138. IEEE, 2020. doi: https://doi.org/10.1109/ACSOS49614.2020.00034.

[18] C. M. Barnes, A. Ekárt, and P. R. Lewis. Beyond Goal-Rationality: Traditional Action Can Reduce Volatility in Socially Situated Agents. *Future Generation Computer Systems*, 113:579–596, 2020. doi: https://doi.org/10.1016/j.future.2020.07.033.

[19] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Behavioural Plasticity Can Help Evolving Agents in Dynamic Environments But at the Cost of Volatility. *ACM Transactions on Autonomous Adaptive Systems*, 2021. In Press.

[20] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Evolving Neuromodulated Controllers in Variable Environments. In *Proceedings of the IEEE 2nd International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, 2021. In Press.

Additional Publications Completed During my PhD Studies:

[14] C. M. Barnes, K. Bellman, J. Botev, A. Diaconescu, L. Esterle, C. Gruhl, C. Landauer, P. R. Lewis, P. R. Nelson, A. Stein, C. Stewart, and S. Tomforde. CHARIOT – Towards a Continuous High-Level Adaptive Runtime Integration Testbed. In *Proceedings of the IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 52–55. IEEE, 2019. doi: https://doi.org/10.1109/FAS-W.2019.00026.

[16] C. M. Barnes, L. Esterle, and J. N. A. Brown. "When You Believe in Things That You Don't Understand": The Effect of Cross-Generational Habits on Self-Improving System Integration. In

Proceedings of the IEEE 4th International Workshops on Foundations and Applications of Self Systems (FAS*W)*, pages 28–31. IEEE, 2019. doi: https://doi.org/10.1109/FAS-W.2019.00020.

[32] J. J. Bird, C. M. Barnes, C. Premebida, A. Ekárt, and D. R. Faria. Country-Level Pandemic Risk and Preparedness Classification Based on COVID-19 Data: A Machine Learning Approach. *PLOS ONE*, 15(10):1–20, 10 2020. doi: https://doi.org/10.1371/journal.pone.0241332.

[96] A. Ghouri, C. M. Barnes, and P. R. Lewis. A Minimal River Crossing Task to Aid the Explainability of Evolutionary Agents. *Artificial Life Conference Proceedings*, pages 36–43, 2020. doi: https://doi.org/10.1162/isal_a_00347.

[21] C. M. Barnes, A. Ghouri, and P. R. Lewis. Explaining Evolutionary Agent-Based Models via Principled Simplification. *Artificial Life*, 27(3), 2021. In Press.

[33] J. J. Bird, C. M. Barnes, L. J. Manso, A. Ekárt, and D. R. Faria. Fruit Quality and Defect Image Classification with Conditional GAN Data Augmentation. *Scientia Horticulturae*, 2021. In Press.

[179] S. G. Robinson, C. M. Barnes, and P. R. Lewis. Centralised and Decentralised Control of Video Game Agents. In *Proceedings of the 20th UK Workshop on Computational Intelligence*, 2021. In Press.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As computer systems become more pervasive in our every day lives with the continual advancement of technology, there are evermore opportunities for these systems to interact with not only humans, but with other systems, components of said systems, and the environments they are situated in as well. These sociotechnical systems can be considered as agent-based, as both human and technical agents make decisions to achieve goals whilst coexisting with other agents that are heterogeneous in their nature. The assumption that modern day computer systems and the components they are comprised will exist in isolation is no longer true. Actors within a shared space exist in an inherently social environment, and can thus interact or interfere with one another [51]; mitigating the consequences of such interference therefore emerges as one of the challenges that must be addressed when designing these socially situated systems.

One contributing factor to the ability of a complex system to function well – specifically sociotechnical systems that are composed of many interacting parts, both human and technical – is said to be resilience to environmental changes [186]. As these systems indeed grow larger, interactions between parts of the system – intended or otherwise, known or not – would increase [101]; it should therefore be possible for these systems to be resilient to, or overcome, such environmental changes brought about by unintended interactions from others. This is especially important because neither humans nor artificial agents possess the information or resources necessary to envisage the consequences of their actions in the far future [171]; if perceiving and reasoning about the consequences of one's own actions is difficult due to resource constraints, then this will also be problematic for understanding

the consequences of the actions of unknown others in the environment. If these systems are not endowed with the ability to respond appropriately to the unknown, there may be catastrophic consequences for integral systems to the safety and functioning of society, such as the emergency services, and traffic management systems [46]. However, since it is infeasible to have complete knowledge of *all* parts of a system because of the sheer processing power required [112] – and that system complexity rises in cases where interference or interdependence is rife amongst such interacting parts [191] – there is a need to move towards systems capable of also overcoming environmental changes as a result of interference when they lack knowledge of its source or consequences.

Possessing the ability to achieve goals when located in a shared, or social environment is said to be a core attribute of 'social intelligence' [88]; the characteristic of 'intelligence' itself is described as a social phenomenon, such that intelligent systems have inherently social qualities that arise from interaction and interference with other systems or parts of systems [51]. Complex behavioural dynamics such as cooperation can therefore emerge from the intentional interaction between different actors within an environment, as well as unintentional, and often unpredictable emergent phenomena that can arise from indirect, and unintended interference from the actions of other actors through the environment itself. Further, in the context of intelligent systems, perceiving and reasoning about models of social concepts, such as the existence of other actors within an environment and the effects that such actors may have on oneself, is described by Bellman et al. [28] as 'social awareness' – an essential component in the endeavour for socially-sensitive sociotechnical systems that can make appropriate decisions to achieve their goals, given that they do not exist in isolation. The ability to learn from and utilise learnt knowledge and experience has been said to be an important component of 'intelligent behaviour' in artificial systems [231]. Designing 'intelligent' systems that are able to tackle new situations – and thus operate in dynamic and uncertain environments with as close-to-optimal performance as possible – is becoming more important as the complexity of the systems and the environments they are situated in increases.

One obvious approach to mitigate this interference might be to extensively model other actors within the environment, such that their actions, interactions, and existence can be reasoned about. However, if no or limited knowledge regarding other actors is available at runtime, this could lead to actors producing inaccurate or incomplete models that result in

potentially misinformed or catastrophic decisions being made [205]. As an alternative to this approach, this thesis instead aims to convey an understanding of how agents can be affected by unintended interactions and interference from other actors, as well as how the consequences of these can be mitigated without such information or modelling. Consequently, this thesis investigates how these complex, heterogeneous real-world systems may begin to overcome the negative effects of interference as a result of existing in an environment shared with many other, potentially unknown actors, as a step towards intelligent, socially-sensitive systems. Agent-based models (ABMs) can be used as a means of understanding the complexities that can emerge in sociotechnical systems, by simplifying or abstracting a real-world problem to study it in detail. This is just one of the uses of ABMs, but they can also be used to identify and resolve potential trade-offs in real-world scenarios, or to spark discussion about certain assumptions in a given domain, for example [145]. Emergent phenomena can be captured in ABMs, which arise from the interactions of actors within a shared environment [34] – much like the sociotechnical systems described thus far. This makes ABMs an appropriate paradigm to study the consequences of interference in order to gain an understanding of, and potentially predict, such consequences in real-world systems. In later chapters, individual agents are evolved in a simulated environment so that they can develop behaviours necessary to achieve their personal goals; the interactions and resulting interference between these agents can therefore be studied.

It is critical that actors within a shared environment – human, machine, artificial agent in a simulated environment, or components of a sociotechnical system – are able to make appropriate decisions, and be resilient to environmental changes despite uncertainty and external changes that can arise from interference from others [210]. Resilience in a system would likely lead to goals being achieved more consistently, due to the system's ability to mitigate the consequences of events beyond its control. In order to understand the consequences of interference within the experimental studies conducted as part of this thesis, the ability of a system to perform consistently needs to also be assessed; this is defined here as the *volatility* of a system, in terms of how much the performance (specifically, the fitness) changes over time. Less volatility would generally indicate a steady performance, and a resilience to any external changes that the system may be experiencing; this would be a desired quality in the socially intelligent systems discussed by Bellman et al. [28] and Castelfranchi [51]. If systems can be endowed with the ability to make socially-oriented

decisions without access to social information, it could mean they are able to tackle the effects of interference whether they are aware of other actors that may be present in the environment or not. By not limiting a system's capabilities to the information it possesses, one would hope that this could lead to better resilience to unknown events caused by others.

1.1 Research Questions

To begin to understand the consequences that interference can have in these complex, heterogeneous real-world systems, a series of experimental studies are presented within this thesis that explore these issues in an abstract agent-based model. In doing this, the way that agents evolve to achieve their goals, when their actions and learning can be influenced by the actions of others, can be studied in detail. Before these systems can express social intelligence, the implications of existing within a shared environment where other actors have the potential to influence or interfere with others needs to be explored; this is so that these systems can be designed with the ability to mitigate any negative consequences that may arise from simply existing in an environment where others also exist. Specifically, this thesis investigates the consequences of interference on agents that possess no ability to gain information about other actors within the environment, or the implications or intentions of their actions; this information may not be immediately available to systems in the real-world. To explore the issues surrounding interference discussed in this chapter thus far, there are three primary questions that this thesis aims to address:

1. How does interference from the actions of others within the environment affect the evolution and goal-achievement of agents, without possessing the ability to perceive or learn of others in their environment?

2. How can the magnitude of the effect that interference has on these agents be measured? Specifically, can this effect be quantified in order to examine how agents experience or mitigate interference?

3. How might agents be designed to mitigate the effects of interference without a reliance on knowledge of others?

To study the effects of interference for Question 1, a simple agent-based model is used to evolve at most one or two agents within an environment to demonstrate the effects

of pursuing goals in isolation (no interference) or in a shared environment (interference). Question 2 is explored by drawing on theory from volatility forecasting in finance, in order to quantify how agents are affected by interference. Finally, social action theory and neuromodulation are operationalised to explore Question 3; taking inspiration from sociology and neuroscience research for the design of agents means that the effectiveness of these two socially-inspired approaches can be explored, in terms of their ability to help agents overcome the effects of interference. The observations and results arising from these experiments can then be compared to establish the effect that interference can have on the ability of agents to achieve their goals, and whether the negative consequences of such interference can be mitigated.

1.2 Contributions of this Thesis

The intention behind the work presented within this thesis is to gain an understanding into how components of real-world sociotechnical systems, as well as artificial systems like ABMs, can be affected by interference – and what the resulting implications are for these systems. As mentioned above, a series of experimental studies have been conducted to explore these questions, using a novel testbed that has been designed for this specific reason. The main contributions of this thesis are therefore as follows:

- A new, gamified, and extensible testbed, the River Crossing Dilemma, that places itself within an established family of testbeds. This is designed to explore how one or many artificial agents evolve to achieve their individual goals in an environment where they may experience interference from the actions of other agents.
- Three metrics – SDoT, CACoT, and CCoT – that can be used to quantify the magnitude of changes in fitness that agents experience during evolution – defined as the *volatility* of evolution. These metrics can be used to analyse the evolution of agents that experience interference in multi-agent environments, as well as those that exist in single-agent environments; the effect of interference can thus be compared in agents that exist alone or are colocated.
- A novel sociologically-inspired approach to determining how agents act within their environment to reach their goals; current systems usually opt to take self-interested

goal-oriented actions, whereas this approach explores how acting in a social manner within a shared environment can affect how agents achieve their goals.

- A novel approach to agent design inspired by theory from the fields of neuroscience and biology, designed to investigate whether temporary, reversible changes in behaviour can be an effective way to mitigate the unpredictable effects of interference within a shared environment.

- Three comprehensive experimental studies and accompanying analyses, designed to investigate the effects of interference, that establish:

  - That agents are able to evolve to achieve their goals when alone in an environment, in line with previous work that explores how agents evolve in isolation. These results act as a baseline for future comparisons;

  - That simply existing in an environment with one other agent whilst pursuing individual goals can affect evolution and goal-achievement, which can introduce volatility that is unpredictable;

  - That changing from a single-agent to a multi-agent environment can affect how agents evolve because of the interference that is introduced, meaning that agents may consequently be unable to achieve their goals because the environment changes in an unanticipated way;

  - That the complexity of the task that agents must complete in order to achieve their goals may affect the level of interference that agents experience, with the effects of interference becoming greater with the complexity of the task;

  - That the predictability of another agent's actions and the resulting interference in a shared environment can influence evolution and goal-achievement; highly variable environments are shown to foster unintentional cooperative behaviour, whereas exploitation is more prevalent in less variable environments.

1.3 Structure of this Thesis

The remainder of this thesis is organised into six chapters with the following structure. The problem of interference in both natural and artificial systems is defined in Chapter 2; a

variety of agent-based models and testbeds designed to explore how many agents exist within a shared environment are identified and evaluated for their appropriateness to investigate the research questions stated above. Additionally, a discussion of how an agent-based model to explore interference can be implemented is presented. Volatility and its implications for artificial systems that share an environment with other actors is also discussed. Three metrics to measure such volatility are introduced in Chapter 3, which are used in each of the experimental studies to analyse the volatility that agents experience during evolution. This chapter introduces a gamified testbed designed specifically to study the effects of interference in artificial agents – one of the major contributions of this thesis. Further, details of the agent design and the evolutionary algorithm used to evolve agents in the experimental studies are also given, as well as an outline of the analytical design for the experimental studies. Chapter 4 introduces the first of the three studies presented within this thesis that explores how interference affects the evolution of artificial agents. Inspired by the way that humans have evolved to act socially in their inherently social environment in order to mitigate interference, social action is operationalised and investigated as a means of mitigating interference in computer systems. The second experimental study is presented in Chapter 5; here, an approach to agent design inspired by theory from neuroscience and biology is utilised as a way to mitigate interference. This study explores how equipping agents with the ability to temporarily and reversibly change their behaviour – known as behavioural plasticity – compares with the widely-adopted goal-oriented approach to agent design in terms of enabling agents to mitigate interference. Additionally, an investigation is conducted into whether the interference that agents experience is affected by the complexity of the task that agents are faced with. The final experimental study is then presented in Chapter 6, which explores the extent to which agent evolution and goal-achievement is affected by the predictability, or variability, of the actions of other agents within the environment. Further, the effect that the variability of the environment – in terms of the changes caused by the actions of other agents – has on the level of interference experienced by agents is also studied, and how this in turn affects how agents are able to evolve to achieve their goals. Chapter 7 then brings this thesis to a close by drawing conclusions from the results of the three experimental studies conducted in Chapters 4, 5 and 6. The research questions and contributions defined above are revisited, concluding with a discussion of the implications of these findings and future avenues of research arising from this thesis.

Chapter 2

Issues of Interference in Natural and Sociotechnical Systems

The work presented in this chapter has been adapted from the following publications:

[15] C. M. Barnes, A. Ekárt, and P. R. Lewis. Social Action in Socially Situated Agents. In *Proceedings of the IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 97–106. IEEE, 2019. doi: `https://doi.org/10.1109/SASO.2019.00021`.

[17] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Coevolutionary Learning of Neuromodulated Controllers for Multi-Stage and Gamified Tasks. In *Proceedings of the IEEE 1st International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 129–138. IEEE, 2020. doi: `https://doi.org/10.1109/ACSOS49614.2020.00034`.

[18] C. M. Barnes, A. Ekárt, and P. R. Lewis. Beyond Goal-Rationality: Traditional Action Can Reduce Volatility in Socially Situated Agents. *Future Generation Computer Systems*, 113:579–596, 2020. doi: `https://doi.org/10.1016/j.future.2020.07.033`.

[19] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Behavioural Plasticity Can Help Evolving Agents in Dynamic Environments But at the Cost of Volatility. *ACM Transactions on Autonomous Adaptive Systems*, 2021. In Press.

[20] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Evolving Neuromodulated Controllers in Variable Environments. In *Proceedings of the IEEE 2nd International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, 2021. In Press.

> "*Since intelligence is mainly a social phenomenon and is due to the necessity of social life, there is the need to construct socially intelligent systems to understand it, and we have to build social entities to have intelligent systems.*" – Castelfranchi [51]

Sociotechnical systems are comprised of many interacting components – be those human, technical system, systems of systems, or a system's constituent parts – where decisions are being increasingly delegated from humans to so-called 'intelligent' machines. These machines typically act in reference to a given goal, such as an objective function, utility or goal-state – which may often conflict with other actors in the system [132]. Yet, humans have evolved both social intelligence and social self-awareness to express more complex behaviour than purely goal-rational action in order to succeed in highly social environments, driven by factors such as values, emotions and traditions [94, 67, 227, 3]. Despite the fact that components in a sociotechnical system are inherently socially situated, they are typically limited to performing goal-rational actions and thus only very rudimentary social action. Specifically, they do not perceive or reason about the effect other systems can have on their own ability to learn and evolve, or the capabilities of others around them; the actions that they direct towards others are also not driven by any broader social meaning – unlike in humans [28, 81]. These systems are also unaware of their *own* impact on the world around them, which can have a catastrophic and unpredictable effect. In 2010 for example, a $1 trillion stock market crash occurred in just 36 minutes, caused at least in part by the unforeseen interactions of several automated trading agents [217].

Sociotechnical systems such as vehicular networks [84], smart energy grids [160], and trading agents [56] are increasingly being designed to operate in dynamic, uncertain and social environments, where interactions are potentially unanticipated or unknown. Explicit and anticipated interactions can be designed for when integrating systems, however, neglecting to consider potentially *unintended* interactions with others that are colocated can lead to worse performance [225]. Hähner et al. [101] argue that as the complexity of these inherently social systems of systems increases, the number of *implicit* and *unintended* interactions between the systems will also increase as a consequence; further, it is not only the interactions with other systems that makes the task of runtime integration challenging, but also as Nelson [159] points out – both the intended and unintended interactions with humans as well. The field of self-improving system integration aims to design systems that

overcome these issues – without full knowledge, control or authority over the other systems in which they coexist and interact with [29]. The actions of one system can have an unintended effect on the others surrounding it [30]; a transition to more socially intelligent systems that are able to learn about others in their environment is therefore necessary to not only enable systems to self-integrate with others around them at runtime, but to maintain learnt knowledge and goal-achieving behaviour despite interference from others.

The remainder of this chapter is organised as follows: the sociality of both natural and artificial systems is discussed in Sections 2.1 and 2.2 respectively, giving an overview of how these types of systems can overcome the challenges associated with coexisting with others; Section 2.3 discusses what 'interference' is in the context of actors that exist in shared environments, as well as the implications that this coexistence can have; various approaches to developing testbeds and agent-based models are outlined in Section 2.4, along with an overview of notable agent-based models that are used to explore how multiple agents exist in shared environments; Section 2.5 then discusses the River Crossing family of environments in detail, which is one family of testbeds that can be used for exploring how agents evolve to achieve goals; further, Section 2.6 explores the importance of measuring the *volatility* of agent evolution, and what this means for comparing how agents are affected by interference; finally, Section 2.7 concludes the chapter by discussing the implications of interference in shared environments, motivating the need to study this phenomenon in detail.

2.1 Evolution of Sociality in Natural Systems

We as humans have evolved the ability to navigate and utilise our social environments to our advantage. Social learning can be seen in both humans and the animal world alike, enabling us to learn and do more complex things than we would be able to individually [219]; it has also been shown to be favoured over individual learning when environmental change is slow, or when individual learning or non-social and environmental cues are not useful [149, 167]. Our innate capacity for cooperation over competition and thus our ability to learn from others has been attributed to our success as a species, and to what distinguishes the complexity of our cognitive abilities from that of primates [91, 153, 105].

The cultural intelligence and cultural brain hypotheses posit that species that have evolved to favour social learning consequently evolve improved asocial learning and individ-

ual problem-solving skills, and flexibility, adaptability and innovation in learning compared to non-social species; this results in a more intelligent population overall [219, 206, 156, 41, 22, 89]. The social brain hypothesis supports this, postulating that social abilities were favoured and acted as a driving force during evolution; this is said to have led to an increase in brain size to support cognitive ability and intelligence, where neocortex size correlates with group size in humans and primates [75, 3, 117, 47].

Kamhi et al. [118] provide supporting evidence for this hypothesis by showing a positive correlation between the size of the mushroom bodies of ants (a higher brain centre like the neocortex in mammalian brains) and colony size. Sociality in ants and other eusocial insects[1] however differs to that of primates; the neural mechanisms driving insect sociality remain relatively unknown [135, 118], but recent efforts explore how this can be addressed in the future [198].

2.2 The Social Nature of Artificial Systems

The success of social learning and behaviour in humans and animals alike has inspired many Computer Science researchers. These observations and theories from nature have been widely used to create optimisation algorithms and to design systems with emergent collective intelligence. For example, self-organisation and cooperation in ants and bees have inspired the development of many optimisation algorithms [71, 119, 53]. Swarm robotics is inspired by the cooperation in groups of social animals to achieve goals or complete tasks, where robots cooperate to solve more complex tasks collectively [43, 154]; this has also been used for societally important tasks such as search and rescue [13]. Social learning strategies have also been widely explored [167, 116], taking inspiration from processes such as mimicry, imitation and learning from others [188, 176, 162].

Individuals in collective systems are inherently socially situated – their actions affect others around them either directly, or indirectly through the environment. Pursuing self-interested action in a social setting can lead to collective irrationality [121]; however, sociality through self-organising institutions can enable groups of self-interested individuals to govern themselves, supporting sustainable management of common pool resources [169]. Social dynamics have been widely explored in areas such as game theory, sociology, economics

[1] Eusociality can be seen in nature in the likes of ants and bees, referring to species that show collective intelligence and where labour is divided amongst groups of sterile workers [118].

and evolutionary computation [218, 212, 121, 131].

Agents may pursue common or individual goals when sharing an environment. Argumentation, negotiation and persuasion [207, 174], goal-aware team affiliation [80], norms and obligations [68], social plans and joint intentions [175, 112], and mutual influence detection [183] all promote cooperative behaviour in multi-agent systems, attempting to mitigate interference between goal-pursuing agents. The BDI (beliefs-desires-intentions) agent architecture for example enables agents to reason about norms and obligations that they are explicitly aware of to promote cooperative behaviour [68]. Other research explores deliberative normative agents, which require explicit knowledge of others, as well as the norms that exist in order to deliberately adhere to or violate them [52].

Sociality is described as the cooperation and organisation of two or more agents in a shared world [51], where agents are goal-oriented, and are social because their actions positively or negatively interfere with one another in terms of achieving goals. Systems that *intentionally* cooperate, coordinate, or act socially, require social awareness [28], and are capable of perceiving and reasoning about others. The evolution of cooperation has been explored extensively, with social dilemmas such as the Prisoner's Dilemma or the Snowdrift Game used to explore social dynamics and strategies [40, 10, 121, 70]. Although the questions that this thesis aims to address surround how agents evolve in environments where cooperation may emerge, *promoting* cooperation is not the focus. Instead, the impact of coexistence and interference on how agents achieve individual goals is explored when they are unable to learn of the existence, goals or intentions of others; knowledge of all others in a system would enable perfect coordination, but is infeasible in dynamic environments due to the infinite power needed for processing and reasoning [112]. Not only this, but information about other systems or actors in dynamic or uncertain environments may not be available at design-time, including their goals, capabilities and how to integrate with them.

2.3 Interference in Sociotechnical Systems

2.3.1 Defining Interference

The term *interference* has been used to describe the interaction between actors in a shared environment, arising from the competition for shared resources [187]. Interference is an inherent characteristic of a shared world; it can arise from actors directly interacting with

each other, or indirectly by interacting with the environment which acts as a passive entity [74, 51]. Existing within a shared environment simply means that any action taken by an agent or actor will affect others [77, 82] – intentionally or not – with the potential to affect agents within a local area, or even globally [107]. The pursuit of goals in a shared world is said to be a core component of *sociality* [51]; social dynamics such as cooperation and competition are the result of interference from actors that help (positive interference) or hinder (negative interference) others with respect to their goals. Other research however states that interference is purely the negative effect on an actor's goal-driven behaviour [144]; here, the term *implicit cooperation* is analogous to positive interference. In this work, the broader definition of interference is adopted, which encompasses both the positive and negative effects discussed above.

Castelfranchi [51] describes *dependence* as a special case of interference, where interference is so strong that actors become dependent on the actions of others to achieve their goals, and thus cannot achieve their goals through their own actions alone. Other researchers have explored the concept of interference with different terminology. Jennings [112] describes this as an *interdependence* of actions that arises when the actions of actors are related, or have an impact on others. Duffy [74] states that co-existing robots will have social interactions that can be indirect, as their actions will *influence* and affect others around them. Rudolph et al. [183] explore how to detect *mutual influences* that arise in smart camera networks, and the resulting dependencies. Thangarajah et al. [211] explore interference *within a single agent*; internal interference can arise from conflicting actions necessary to fulfil parallel conflicting goals (e.g. an action is a step towards achieving one goal, but a step away from achieving another, conflicting goal).

These methods require agents to be aware of others in the environment, and potentially their goals or intentions. In unpredictable and dynamic environments, this is not always possible; the environment and the agents within it may change over time, so this information may not always be available or predictable at design-time. Tomforde et al. [215] identify that subsystems can influence one another either directly or indirectly, whether they are intended to interact together or not; consequently, the 'interwoven' nature of these systems where uncertainty, heterogeneity of actors and interference between said actors prevail means that traditional methods of integration become impractical [27]. Thus, agents in dynamic or uncertain environments, where the presence of others is potentially unknown, need

the ability to develop and maintain these models on-the-fly. Endowing systems with the ability to learn from their inherently social environment, as human societies have evolved to do, could therefore enable them to learn and evolve potentially more complex individual behaviours, problem-solving competencies, and goal-achieving abilities.

2.3.2 Implications of Interference

The technical systems of today are growing in size and complexity, and thus both interactions and interference between components and through the environment are evermore prevalent [101]. Aldelaimi et al. [5] highlight that as the number of devices in complex systems such as smart cities increases – as well as the heterogeneity of such devices – perceiving, learning about, and interacting with other devices is challenging without human intervention. Burger et al. [46] go further to note that it is increasingly important for the devices involved in emergency response services, waste water treatment, traffic management and even Wi-Fi – all critical for the functioning and safety of smart cities – to be resilient, in order to prevent and avoid high work loads or catastrophic loss. As the complexity of these large, heterogeneous systems increases, unknown or unforeseen situations due to interference caused by the actions of others will be encountered more often. This must be dealt with appropriately for the safety and functionality of both the systems as a whole and the environments in which they operate. Since information about others, or the ability to learn of others, within the environment may not be immediately available – or at all – in a dynamic or complex environment [112], an investigation into how interference affects actors in an environment without this knowledge needs to be conducted.

2.4 Studying Interference using Agent-Based Models

Various toolkits and approaches to developing agent-based models and simulations exist, each with their own application domains, benefits and drawbacks. The research questions detailed in Chapter 1 fundamentally aim to observe how agents may evolve to achieve their goals when their capabilities (in terms of perceiving and reasoning about others, and their abilities, goals and intentions) are limited, and how the presence of other agents may affect this. In this section, a brief overview of some common agent toolkits and approaches to agent design is given, to motivate the approach adopted when developing the experimental

studies in later chapters of this thesis. Further, existing agent-based models that are used to investigate how multiple agents exist within an environment are discussed, to assess their suitability for exploring the consequences of interference.

2.4.1 Agent Toolkits, Frameworks and Design Approaches

The 'Multi-Agent Simulator Of Neighborhoods' (MASON) agent toolkit [137, 138, 139] is a platform-independent, extensible simulation library; this was developed to help researchers create and visualise simulation experiments that potentially comprise large numbers of agents and interactions. MASON has been used in various applications, including those relating to the modelling of complex societal issues, for example: to explore the evolution of cooperation when the population size can vary [193]; in collaborative foraging tasks [108]; to simulate the effectiveness of different disaster response mechanisms, positioning of aid centres, and the impact of socio-cultural information surrounding relief efforts [60]; and has been extended for example to the MASON RebeLand model, in order to explore political and societal issues in relation to governmental performance [55]. Despite the adoption of MASON in a number of domains, according to Kravari and Bassiliades [126] however, MASON has a complicated interface which requires developers to overcome a learning curve before using. Since the research questions outlined in Chapter 1 concern only a few agents, MASON is not considered appropriate for the experimental studies in this thesis.

Whilst the MASON toolkit is proven to be extensible and applicable to a wide range of research domains, Kubera et al. [128] highlight that there is a high probability that developers would introduce errors when utilising the toolkit, as there are no guidelines or constraints regarding the design of agents, or their behaviour or interactions. To address this issue, Kubera et al. [128] outline the 'Interaction-Oriented Design of Agent simulations' (IODA) approach, which includes a methodology and model for agent design, and the 'Java Environment for the Design of agent Interaction' simulation framework (JEDI) to create the simulations. Being designed around the principle that interactions are composed of actions that involve one agent, and either another agent or the environment, the IODA approach would facilitate the exploration of the research questions defined in Chapter 1: how the consequences of inter/actions with the environment may interfere with or affect other agents. However, the authors state that this approach is not intended to be used for simulations where complexity does not arise from the *number* of interactions between the agents in

the simulation; for exploring the effect of interference where agent interactions cannot be intentional or direct, this approach is therefore unsuitable for designing experiments to explore the research questions stated in Chapter 1.

JADE, which stands for 'Java Agent DEvelopment Framework', is a FIPA-compliant[2], open-source agent platform and development framework implemented in Java, designed for the development of multi-agent systems [26]. Whilst JADE is a popular choice for researchers in both academia and industry [126] – which can be partially attributed to its learnability – one concern about its suitability for the intentions laid out in Chapter 1 regards agent communication. For example, an extension of this framework, Jadex (JADE eXtension) [6], incorporates the BDI architecture (Belief-Desire-Intention, see Section 2.2) for the design of goal-oriented, rational agents with the capacity to communicate with one another. The intention of this work however is to explore how agents are able to pursue their own goals, *without* knowing of, or being capable of communicating with, other agents; as such, this is unsuitable for exploring the questions posed in this thesis.

One of the most popular ways to implement multi-agent environments or to model complex phenomena is with NetLogo, an integrated modelling environment that is written in Java and designed for use in both research and teaching [233, 213]. NetLogo can even run in the browser, with numerous sample models and code examples available to run in the areas of art, biology, computer science, mathematics, and more[3]; this makes it simple to start experimenting with pre-set models while learning how to navigate the user interface to customise the models themselves. NetLogo itself is also a programming language used to set up the experiments or simulations in the modelling environment, so there is some learning involved. The developers have also worked to increase the extensibility of NetLogo, allowing users to extend, replace or add components in Java using APIs. Despite this, similar concerns to those that have been voiced about the other approaches discussed above apply; care would need to be taken when extending NetLogo to run custom simulations, and a good understanding of how to use the APIs themselves to do such a thing is crucial.

[2]The Foundation of Intelligent Physical Agents (FIPA) is an IEEE Computer Society standards organisation: `http://www.fipa.org/`.

[3]NetLogo Web is available at: `http://www.netlogoweb.org/launch#NewModel`.

2.4.2 Exploring Interference with Agent-Based Models

Agent-based models (ABMs) have been used to simulate and explore complex real-world problems in a number of research areas, where a real-world scenario or natural phenomenon is simplified in order to study it in closer detail. This indeed is one of the aims of the Artificial Life research area – to understand or gain insight into the living world and the complex phenomena that emerge [25, 4, 96]. ABMs have consequently been used to explore real-world issues such as the management of water resources in sociotechnical water systems [31], to analyse social patterns that emerge from collections of individuals [166], and to investigate cooperative path planning strategies in autonomous guided vehicles [140], to name just a few examples. Since interference is a complex phenomenon that can emerge from the indirect interactions between actors and the environment, an ABM can be used to explore the consequences of interference in fine detail by using a scenario with reduced complexity. The research questions defined in Chapter 1 however are formulated around understanding both the consequences of interference, and how this can be mitigated by actors (either agents in a simulation, components of a system, or an entire system itself) that may not have the required knowledge about others – or ability to acquire this knowledge – to make appropriate decisions whilst experiencing interference. With this in mind, an appropriate model or testbed needs to be selected in order to investigate these issues that arise in agent-based systems.

Many testbeds already exist that have the capacity to explore how multiple agents act within an environment, such as: the MICE testbed (Michigan Intelligent Coordination Experiment) [76], designed to explore coordination and interactions between agents in a 2D world; MAGES (Multi AGEnt System) [38], designed to observe different interactions between heterogeneous agents; and Dedale [104], designed to study the learning and decision-making processes of agents that may coordinate to solve a task; Evosphere, which is used to study the evolution of populations of 3D creatures with natural selection that inhabit a 'micro-planet' [151]; Polyworld, designed to explore foraging behaviour in agents controlled by neural networks that exist in a 2D world [235]. Whilst interference is not the focus of the experiments conducted using these testbeds, they can all simulate multiple agents existing within a shared environment – so interference can emerge from the actions of each agent. Other studies explore interference when agents possess an explicit

knowledge of others; this knowledge is then used by the agent to make informed decisions regarding the interference, and their goals and action choices. Rudolph et al. [183] for example simulate a smart camera network with multiple cameras, in order to explore how 'mutual influences' (i.e. interference, see Section 2.3.1) can be detected, and thus mitigated using knowledge of others. Godoy et al. [97] propose a distributed coordination approach for agents where the actions or motions of others around them can negatively affect path planning and goal-achievement; this decentralised approach to mitigate the interference that agents experience utilises knowledge of others within a certain range, and one-way communication between agents in order to avoid collisions. Michael et al. [150] on the other hand, simulate heterogeneous multi-robot systems in order to validate the observations in real-world robots; there is a disparity between simulation and reality when the actions and interactions of each robot may interfere with others, which is mitigated by endowing the robots with the ability to sense and communicate with others. Malakuti [141] explicitly aim to detect unexpected and undesirable emergent interference when integrating multiple self-adaptive systems together; however, the authors state that this detection requires extensive knowledge of each individual in the system and the consequences of the interference. This would quickly become infeasible as the scale of the system increases, due to the number of interacting components [112, 107, 101].

Interference can arise in any scenario in which there is more than one actor within an environment, and as such, interference is not a new concept when talking about ABMs with more than one agent – as discussed in Section 2.3.1. However, whilst interference is an emergent property in ABMs with multiple agents, the focus of the studies that use these multi-agent ABMs is usually on: mitigating or detecting the interference that arises *when agents have knowledge of others* [141, 183]; or developing strategies to facilitate the emergence of social dynamics such as cooperation, competition, or coordination for example, where interference is an implicit property of the simulation due to its multi-agent nature but is not explored [97, 104]. The effect of interference can potentially increase in severity like a ripple effect, initially having a local effect but can continue to affect the system globally [107] when there are many agents within an environment. The actions of one agent can thus have a knock-on effect on other agents, which in turn affect other agents, and so on. In each of the examples discussed above, the studies conducted observe the aggregate behaviour of many agents. In order to understand the consequences of interference at a fundamental level

however, there is a need for a simple testbed that can be used to observe and analyse the differences between existing alone (i.e. no interference) and existing with *one* other agent. By reducing the complexity of the environment from including many agents to simply one or two, the potential magnitude of the interference experienced by agents is also reduced as a result. None of the testbeds or ABMs above are therefore adequate to observe the effects of interference in close detail, as each of the examples discussed are made complex by the sheer numbers of agents in the environment.

2.5 The River Crossing Family of Environments

Following the discussion in the previous section about existing testbeds to explore interference, this section introduces a family of environments typically used to study how *one* agent evolves in an environment to solve tasks. Since a suitable testbed that can be used to study the effects of interference in close detail has not yet been identified, modifying an established testbed could facilitate this investigation instead.

2.5.1 The River Crossing Task

The River Crossing Task (RC Task, or RCT) was developed by Robinson et al. [178], in order to explore how agents can evolve to solve tasks in dynamic environments. They introduce a novel, two-tiered neural network architecture that enables artificial agents to express both deliberative and reactive behaviours, giving them the ability to navigate dynamic environments without the need for *a priori* knowledge of the task or environment; these behaviours are acquired over the course of evolution, where a population of individual agents is maintained. The goal of agents is simple: they must collect the Resource object to receive a reward of a highly positive fitness, which is located on the opposite side of a river that spans the length of the environment. The RCT testbed itself is a simple 2D grid, containing objects such as Traps, Stones, and a river consisting of Water objects. The difficulty of the task arises from the fact that agents must learn 'sub-goals' to retrieve their reward object (the Resource), such as building a bridge in the river with Stones; this is challenging as falling into the river gives the agent a highly negative fitness, so learning the goal-achieving behaviour of building a bridge is initially risky. The complexity and dynamicity of the environment varies within the study; agents are evaluated on three

consecutive environments of increasing difficulty at each generation, with the width of the river increasing from a width of one (firstly where bridges already exist, and secondly where no bridges exist) to two cells. Further, Robinson et al. [178] concretely demonstrate that agents are able to react immediately to environmental changes by evaluating agents in environments where walls of Traps move locations, for example.

2.5.2 The RC Family

The simplicity and ease of extensibility of the RCT inspired the creation of other testbeds, thus leading to the development of what is termed here a 'family' of River Crossing (RC) environments; these all have a common learning challenge – agents must learn to build a bridge in order to achieve their goal. Borg et al. [37] for example introduce a more complex version of the RC task called the RC+ task; this is used to demonstrate that learning by imitation through transcription errors and cultural transmission can enable agents to achieve goals where incremental evolution cannot. The RC+ task is therefore specifically designed such that agents are evaluated on five consecutive environments at each generation, where the final environment cannot be solved by incremental evolution on its own. In this study, the width of the river increases with every evaluation from zero to four, whilst the number of Stones available to build bridges reduces to zero; consequently, the task in the final environment can only be solved by placing a new, unseen object into the river. Agents that use a teacher-learner strategy with transcription errors are shown to be capable of solving this final, complex environment; agents that are 'non-learners' on the other hand are unable to solve this problem, even with the addition of transcription errors or a higher mutation rate.

A further adaptation of the original RC task is presented by Stanton and Channon [203], termed the 3D River Crossing task (3D RC), which is later extended into the Physical 3D RC task (P3D RC) [204]. The 3D RC is used to explore how 3D quadruped agents are able to evolve reactive and deliberative behaviours, whilst navigating a world that requires attention to the locomotion of the agent's limbs in order to achieve the goal. The P3D RC develops this further by also requiring the 3D agents to manipulate objects directly within the environment using physical motor control, instead of the abstract interaction which is seen in the 3D RC.

A more recent addition to this family of environments is the Minimal River Crossing task

(RC–), presented by Ghouri et al.[4] [96]. The intention of this version of the testbed is not to observe how agents are able to express reactive or deliberative behaviours, but instead to explain and understand agent *why* the core learning challenge in the River Crossing family of environments – building a bridge – is in fact challenging. The original RC task [178] is reduced from a 20×20 grid to a 1×5 grid, such that this core learning challenge can be isolated and observed in fine detail; one of the main contributions of this study is that the RC– can be used to predict and explain agent behaviour that generalises back to the original problem.

2.5.3 Studies Conducted in the RC Family

In addition to the studies that first introduced these instances that make up the River Crossing family of environments, both the testbeds themselves and the agent architecture used to evolve these agents have been further modified in other work to explore a broader range of concepts. For example, the 3D virtual creatures designed by Stanton and Channon [203] extend the two-tiered learning architecture first introduced by Robinson et al. [178], to observe how 3D agents may navigate the 3D RC environment; the addition of new physical and pattern generator networks to the initial two-tiered architecture facilitates the evolution of locomotion, turning and avoidance behaviour in 3D agents.

The RC+ task [37] on the other hand has been used by Jolley et al. [116] to investigate the effect that the choice of 'teacher' in different teacher-learner social learning strategies can have on goal-achievement and evolution; this develops upon the study presented by Borg et al. [37], which exclusively assigns the winner of the last tournament to be the teacher, and the aforementioned agent's most recent offspring to be the learner (known as the 'best parent' strategy). Jolley et al. [116] show that the choice of social learning strategy has no significant effect on fitness when comparing a teacher that is the best parent, the fittest, the oldest, the youngest, or a random agent; each strategy is effective for enabling agents to solve tasks using behaviours which are inaccessible to incremental evolution alone.

In other work, Jolley and Channon [114] use the RC task to demonstrate an alternative to the two-tiered architecture proposed by Robinson et al. [178]. The second of these two tiers is a topologically-organised neural network responsible for reactive behaviours and motion

[4]The author of this thesis co-supervised the first author of this publication [96], who studied the explainability of agents for her BSc dissertation project, by introducing the Minimal River Crossing Task testbed (RC–).

planning, however this is replaced by a HyperNEAT [201] implementation to compare the effectiveness of the approach to the original. Introducing a modified fitness function is shown to improve the general performance of agents using this adapted agent architecture; agents are evaluated on multiple RC instances to simulate learning in a 'large' environment, and as such appear to learn more generalised behaviour. Despite this, the new HyperNEAT implementation of the RC agent architecture is not shown to improve agent performance compared to agents operating in the original RCT. In later work [115], the second tier in the RC agent architecture is instead replaced with a single-layer convolutional neural network (ConvNet) [129]; this is shown to significantly improve agent performance in the RC task compared to the HyperNEAT implementation.

The use of the agent architecture introduced by Robinson et al. [178] is not exclusively used in agents that operate in instances of the River Crossing family of environments. Borg and Channon [35] demonstrate that this agent architecture is effective for evolving agents in the EnVar environment; agents must forage for resources or 'plants', aiming to consume those that give positive rather than negative energy values. The focus of this study is to observe whether social information affects the evolution and adaptivity of agents, concluding that a benefit is indeed seen consistently when the difficulty of the task is low. Task difficulty here is in terms of the ratio of positive to negative food resources available in the environment, where difficulty increases with the number of negative food resources. Social information about an agent that is being perceived however can be simply regarding its presence, the action it is currently performing, its health or energy level, its age – or alternatively where no social information is perceivable; agents have no control over the social information that others can obtain from them, as this transfer is dependent on the type of information that an agent can perceive. This is emphasised in later work [36], where these simple forms of social information are observed to promote interaction between agents, where movement towards other agents to obtain more social information can lead to improved performance.

2.5.4 The Bigger Picture of Crossing Rivers

In general terms, learning to cross a river in a simulated environment is not an important task, nor is it an immediately useful exercise. What *is* important and useful however, is what can be learnt from how agents may interact with, and evolve within this simplified en-

vironment. Observing how simple agents may respond to different environmental conditions in abstract terms therefore has the potential to lead to conclusions that can influence and benefit how future technical systems should be designed, when facing similar conditions. The River Crossing family of environments and the studies which utilise them are firmly placed in the realm of Artificial Life research – an aim of which is to use simplified models to study and understand complex natural phenomena [96, 4, 25].

The act of learning the behaviours necessary to cross a river seems simple, but completing such a task has been proven to be difficult using this well-established family of testbeds. In the River Crossing task, agents must learn two implicit, and conflicting concepts about the environment: the first is that the river should be avoided at all costs due to the risk of drowning; secondly, interacting with the river is not only safe to do *under the condition that* a Stone is being carried, but is in fact *necessary* in order to achieve the goal. The River Crossing task is thus an abstraction of some of the complex challenges that technical systems may face in the real world; designing systems to achieve individual goals in dynamic environments is challenging, and will be increasingly so as the size and complexity of these systems, and their resulting interactions, also increases [101]. Observing how artificial agents develop the ability to perform sub-tasks to achieve goals in the RC family can therefore aid understanding of these issues in simplified terms.

An attractive characteristic that is common throughout the River Crossing family of testbeds is the ease of tuning the difficulty of the task within the environment, depending on the focus of the study. Further, these testbeds have been used to explore a diverse range of concepts, such as the effects of social learning and social information, the ability to evolve robust behaviour in dynamic environments, and even the impact that the type of agent learning architecture itself can have on deliberation and goal-achievement. As discussed above, previous studies conducted using the River Crossing family of environments have explored how agents evolve in dynamic environments. This dynamicity arises from novel and unseen environmental configurations or moving objects, meaning that it is impossible to follow a predetermined or planned route to achieve the goal. However, it is becoming commonplace that the environments that real-world systems operate in can also change, or become dynamic, due to the actions of other systems or actors [191, 101, 210] – be those human, or artificial (as is the nature of sociotechnical systems in general). An environment inhabited by more than one individual is inherently social, with the actions of

others directly or indirectly interfering with how individuals perceive and interact with their environment [51]. Bellman et al. [28] state that systems operating in shared environments should express *social awareness* in order to perceive and learn about the existence of others in their surroundings, as well as their capabilities, goals, and even intentions. However, the ability to perceive other systems, as well as how to interact with them, may not be immediately accessible [112] – an issue that is only worsening as the heterogeneity of today's sociotechnical systems increases. The questions posed in this thesis are therefore designed to provide some understanding of these issues, and how they may be overcome.

2.6 Volatility and its Implications in Sociotechnical Systems

The performance of evolutionary algorithms is often analysed by ascertaining the expected time taken to find an optimal solution in terms of received fitness; this can then be used to compare the ability of an algorithm to perform with varying parameters or conditions, or to compare the performance of the algorithm to alternative algorithms. Jansen and Zarges [111] note that optimisation algorithms that operate in static environments are intended to find solutions as quickly as possible, however this changes in dynamic environments as the optimal solution may change over time due to varying environmental conditions.

It is important to clarify terminology when discussing algorithms that operate in dynamic or shared environments, as to not conflate different areas of research. Evolutionary dynamic optimisation problems – that is, optimising evolutionary algorithms in dynamic environments [92, 161, 113] – are characterised by optimising solutions when the optimum may change over time [92]. One common way to analyse the performance of optimisation algorithms is to measure the best-in-generation fitness, either averaged over all generations, or for each generation over multiple runs; comparing the performances of different algorithms with these measures can however be difficult if they are not normalised [161]. A separate, but related, concept to the dynamic optimisation problems described above is coevolution, where multiple individuals or species are evolved to solve optimisation problems [165, 239, 243]. In coevolutionary algorithms, an individual's fitness is *dependent* on the fitness of others from either the same population, or a different one [243, 72]; the fitnesses of individuals therefore interact or are coupled, which distinguishes *coevolutionary* from *evolutionary* algorithms [85, 182]. Coevolution can be competitive, cooperative, or

even both [243], where an individual's fitness is either negatively or positively affected by the success of another respectively. The studies presented within this thesis concern how one agent evolves to achieve an individual goal when it may share an environment with, and be affected by unknown others; as such, analysis is conducted from the perspective of an individual agent, rather than the dynamics observed in all individuals. While beyond the scope of this thesis, analyses in a coevolutionary or other sense may be conducted in the future, in order to understand the evolutionary dynamics that emerge in these experimental studies further.

2.6.1 Analysing Volatility

The experimental studies introduced in later chapters may evolve populations of agents in either static or dynamic environments (where dynamicity is caused by the actions of others within the environment, rather than from the environment changing of its own accord or by the goal itself changing over time), in order to explore how interference from other, unknown actors within the environment may affect evolution and learnt behaviour. Experiencing dynamicity presents a challenge when analysing the performance of algorithms, because an optimal solution may become sub-optimal if conditions change. Simply capturing the number of generations (i.e. the time taken) for an evolutionary algorithm to find a successful solution is therefore inadequate to assess the algorithm's suitability to the environment, since the ability to mitigate interference may fluctuate during evolution depending on environmental stimuli. As an alternative to this approach, the algorithms used in the experimental studies are instead analysed over the entirety of evolution; to assess the extent to which the populations of agents in the later studies are affected by interference (and thus dynamicity within the environment), the fitness received by agents is tracked over evolution. One would expect that the more the fitness changes during evolution, the more the algorithm is susceptible to the negative (or positive) effects of interference. A fitness that changes often from one generation to the next would consequently indicate that it is hard for the agents in the population to maintain a constant fitness.

It is important to define terminology – especially when conducting inter-disciplinary work – not only to avoid ambiguity and aid understanding, but because terms may have different meanings depending on both the context or field of study in which they are used. In plain English, one could describe the extent to which the fitness of a population is maintained

during evolution – as discussed in the previous paragraph – as its 'stability'. However, this is a widely used term that is usually associated with the exploration of 'evolutionarily stable strategies' [194] or 'evolutionary stable states' [195] in the fields of game theory [194, 86, 134], economics [24], and biology [194, 39]. In these areas, 'stability' refers to the stability of the strategies that are employed by the population, such that the population itself exists in a state of equilibrium and is thus robust to the intrusion of mutant strategies. Similarly, 'stability' has also been used to describe learning dynamics in multi-agent reinforcement learning scenarios, referring to agents that converge to an equilibrium or stable policy, such as a Nash equilibrium [44]. Since this term already has a specific meaning in a number of different fields of study, it would be illogical to introduce an alternative meaning in this thesis to describe the fitness received by agents during evolution. Instead, an already established term that has a similar meaning is 'volatility', which is used in the area of finance to model or forecast the volatility of financial markets [170]. A higher volatility means that there is a higher variability and dispersion of the values [234] – volatility increases in line with the number of fluctuations detected in the financial model. This can be easily translated to the domain of evolutionary agent-based models, since capturing the fitness received by agents over the course of evolution is similar to tracking the fluctuations present in a financial model. One of the simplest, and most common, ways to quantify volatility in finance is to calculate the sample standard deviation over time [170]. It must be noted that the use of the term 'volatility' here is unlike that of Vega-Redondo [221], which uses the phrase 'equilibrium volatility' in the context of evolutionary game theory to describe the dynamics of equilibria during evolution – without quantifying such volatility. Taking inspiration from this method, volatility can hence be used to quantify and describe the fluctuations in fitness during the evolution of agent-based models; this approach is grounded in theory that can be applied in the context of evolutionary algorithms, without introducing ambiguity – which would be a consequence of adopting a term such as 'stability'.

Koren and Tenreyro [124] highlight that GDP growth is more *volatile* in poorer countries compared to rich countries, caused in part by weaker financial infrastructure that struggles to deal with economic shocks or political instability appropriately, for example. Without the means to combat such unpredictable events, growth rates in these countries fluctuate more often, and by larger amounts. This is analogous to simulated agents that are unable to respond appropriately to environmental stimuli during the course of evolution. Agents

that evolve in dynamic environments, that have not yet evolved behaviour that is robust to unpredictable or unknown events, will receive a fitness that often fluctuates; this would result in agents that experience *volatile evolution*. Section 3.6.1 introduces three metrics which have been devised to quantify the amount of volatility that agents experience during evolution, by observing how the fitness changes over time. By using these metrics to analyse the performance of evolutionary agents, or indeed sociotechnical systems situated in the real world, one would expect that a lower volatility would imply that the system or agent is better equipped to deal with unforeseen circumstances, and is able to maintain a more constant fitness or level of performance compared to more volatile actors. This being said, low or near-zero volatility may also be an artefact of a system that is unable to achieve its goals, and consequently maintains a constant level of sub-optimal performance. As such, the volatility metrics described in Section 3.6.1 should therefore be used in combination with an analysis of the actual performance in order to avoid any incorrect conclusions being drawn (in this case, an analysis of the fitness received by the agents in the experimental studies in later chapters is used in conjunction with an analysis using the three volatility metrics defined). Once an analysis of volatility has been conducted, the evolution of different populations of agents can consequently be directly compared. This comparison can then be used to indicate whether one approach to agent design is preferable over another (in terms of the amount of evolutionary volatility experienced, and thus the ability of an agent to maintain a constant fitness), or to demonstrate the effect of various environmental stimuli on evolution – such as the effect of interference as a result of sharing an environment with another agent, compared to existing in an environment alone.

2.7 Conclusion

The research questions that this thesis aims to address surround understanding *how* systems may affect one another when coexisting with others in a shared space, and consequently how systems may evolve to achieve their goals despite possessing no ability to perceive or learn of others. The investigation of whether, and how, simulated agents may achieve their individual goals under these conditions is considered here to be an important prerequisite to the 'social awareness' that is discussed by Bellman et al. [28]. One of the aims of this thesis is therefore to investigate whether a system is able to mitigate the effects of interference

before any information can be learnt about others, since this information or the ability to acquire it may not be immediately obtainable in the real-world.

The versatility of the River Crossing family of environments has been demonstrated through numerous studies and extensions of the original testbed itself, thus demonstrating the viability of the River Crossing task for exploring complex problems in abstract terms. A limitation of this family of environments however is that there is a lack of capacity for exploring the effects of interference in shared environments. Each of the studies conducted to date concern how artificial agents evolve when they are the sole inhabitant of the environment; therefore, an addition to this family of environments that facilitates the exploration of interference and its consequences is needed, to understand how agents may evolve to pursue their *individual* goals when resources in the environment are shared with others. The design of this new testbed, intended to facilitate the exploration of interference and volatility, is presented and discussed in detail in Chapter 3.

In this chapter, it is established that interference can arise in environments shared with multiple actors – be those human, machine, or a combination of the two. Current approaches to mitigating interference require information about others in order to make decisions, but this might not always be possible – especially since the size and complexity of real-world computer systems is increasing. Consequently, the effect that this interference can have on actors in a shared environment needs to be understood, which can be achieved by observing how actors (specifically artificial agents) evolve in a shared environment without such knowledge. By exploring the consequences of interference, one may begin to think of how systems could be designed to mitigate interference without having to rely on complete or accurate knowledge of others. This will be increasingly important to consider when designing real-world sociotechnical systems, since the interactions between parts of the system, and therefore the effects of interference, will scale with the system size. Agent-based models are widely used to understand complex real-world phenomena in a simplified manner, meaning that this is also appropriate for exploring the implications of interference. An appropriate testbed is required to conduct such an investigation, as well as a means of measuring the effect that interference has on agents during evolution. The insight gained from observing how artificial agents evolve in shared environments when experiencing interference can consequently be used to make informed decisions when designing real-world sociotechnical systems that experience interference on a larger scale.

Chapter 3

A Testbed to Explore Interference in Agent-Based Systems

The work presented in this chapter has been adapted from the following publications:

[15] C. M. Barnes, A. Ekárt, and P. R. Lewis. Social Action in Socially Situated Agents. In *Proceedings of the IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 97–106. IEEE, 2019. doi: https://doi.org/10.1109/SASO.2019.00021.

[17] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Coevolutionary Learning of Neuromodulated Controllers for Multi-Stage and Gamified Tasks. In *Proceedings of the IEEE 1st International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 129–138. IEEE, 2020. doi: https://doi.org/10.1109/ACSOS49614.2020.00034.

[18] C. M. Barnes, A. Ekárt, and P. R. Lewis. Beyond Goal-Rationality: Traditional Action Can Reduce Volatility in Socially Situated Agents. *Future Generation Computer Systems*, 113:579–596, 2020. doi: https://doi.org/10.1016/j.future.2020.07.033.

[19] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Behavioural Plasticity Can Help Evolving Agents in Dynamic Environments But at the Cost of Volatility. *ACM Transactions on Autonomous Adaptive Systems*, 2021. In Press.

[20] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Evolving Neuromodulated Controllers in Variable Environments. In *Proceedings of the IEEE 2nd International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, 2021. In Press.

3.1 Motivation

As discussed in the previous chapter, agent-based models are widely used as a way to simplify a real-world scenario in order to study it in detail; consequently, a simple testbed can be used to study the implications that interference can have in sociotechnical systems, by reducing the complexity of the scenario and studying the fundamental issues that arise for artificial agents in a shared environment. In Section 2.5, the River Crossing (RC) task is identified as a well-established testbed for exploring how agents evolve to solve tasks in complex, and dynamic environments; the integral learning challenge incorporated into this testbed (the act of building a bridge to achieve a goal) has inspired a variety of extensions to the original RC task to explore different phenomena in both 2D and 3D simulated agents. The RC environment itself is simple, lending itself well to extension and further development for use in different domains; consequently, a testbed designed to observe how multiple agents may achieve their goals in this environment seems both fitting, and a valuable addition to this family that has been growing in popularity in recent years.

The purpose of this chapter is to introduce the River Crossing Dilemma (RCD); this is an agent-based model designed to explore how artificial agents evolve and behave in shared environments. Specifically, the testbed builds upon the original River Crossing task testbed proposed by Robinson et al. [178], in order to investigate the effect that interference arising from the actions of others can have on the evolution and fitness of agents. The importance of understanding interference and the way it can impact a system's performance or ability to achieve goals is becoming increasingly important, due to the increasing size, complexity, and heterogeneity of today's sociotechnical systems. This testbed can be used to both understand and explore the implications of this interference by observing how artificial agents react to, and potentially mitigate, interference from others; as such, the RCD can be used to study what the implications of interference are for agent-based systems in the real-world. To simulate the changing and uncertain conditions that sociotechnical systems increasingly face, the agents operating in the testbed discussed in this chapter (as well as the studies in later chapters that use this testbed) do not possess the ability to learn of or perceive others within the environment. Information about others may not be immediately available in the real-world; exploring how systems can be designed to be resilient to the effects of interference without being aware of its cause or existence is therefore hypothesised

to become more critical as the number of interactions between systems increases [101].

The remainder of this chapter is organised as follows: the River Crossing Dilemma testbed is introduced and discussed in depth in Section 3.2, while Section 3.3 discusses agent design; Section 3.4 details the evolutionary algorithm used to evolve agents in the testbed; the adopted approach to implementing a testbed is discussed in Section 3.5; Section 3.6 outlines the approach for analysis used in each experimental study, with Section 3.6.1 specifically introducing three metrics that can be used to measure evolutionary volatility; Section 3.7 finally concludes by outlining the contributions of this chapter.

3.2 The River Crossing Dilemma Testbed

One of the contributions of this thesis is the River Crossing Dilemma (RCD) testbed[1], a gamified testbed designed to explore how agents evolve to achieve individual goals in shared worlds. This is an extension of the original River Crossing task proposed by Robinson et al. [178], where the gamification of the RCD environment introduces the opportunity to study how multiple agents act in inherently social situations. These may be tractable social dilemmas, such as those introduced in the studies in later chapters, but in general are not constrained in their complexity since RCD instances may be designed to be arbitrarily complex. Agents have no prior knowledge of the task or environment, and must learn what their goal is and how to achieve it without this information. Agents pursue their own, individual goals, and do not require knowledge of others to do so in shared environments, as they have the potential to perform the actions necessary to achieve their goal solely relying on their own behaviour.

The RCD is a 19×19 grid-world environment with a two-cell deep river of Water in the centre (Figure 3.1). There are four Stones on each river bank, and all empty cells are Grass. As this testbed is designed to study how interference affects evolution and goal-achievement in agents, the RCD reduces the variety of objects in the environment compared to the original River Crossing (RC) task introduced by Robinson et al. [178]; objects such as Traps, which are seen in the RC task, increase the complexity of the task at hand without contributing to the study of interference. Agents therefore only encounter the objects that are sufficient to achieve their goals.

[1]The River Crossing Dilemma testbed was first published in Barnes et al. [15].

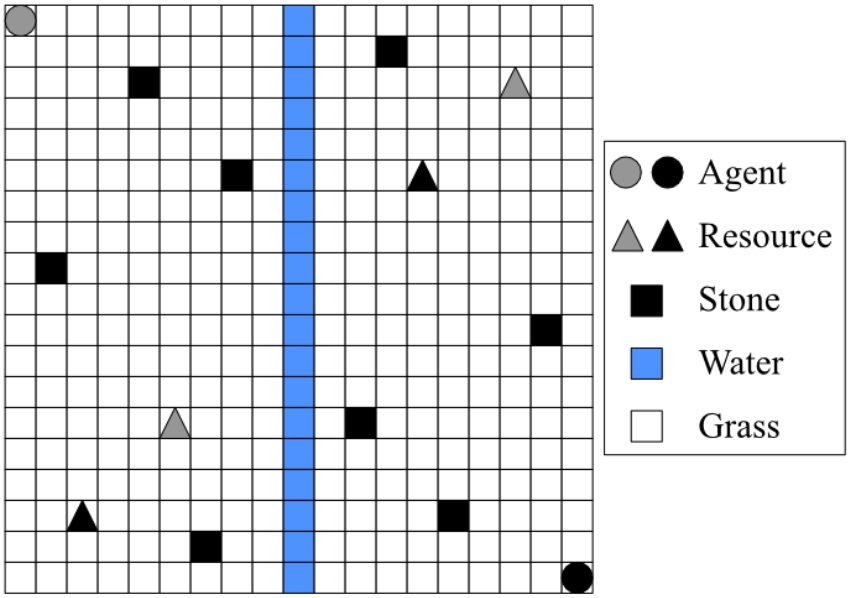


Figure 3.1: The River Crossing Dilemma testbed, which is a 2D grid-world environment. The grey agent (top left) is allocated the two Resources in grey, and the black agent (bottom right) is allocated the two Resources in black; agents cannot interact with Resources that are not allocated to them. Both agents can interact with all other objects. For single-agent environments, the black agent is removed.

The final type of object in the RCD environment is the Resource. An agent's individual goal is to collect both of its two allocated Resources, which are placed one on each side of the river; achieving the goal will reward the agent with a highly positive fitness. Conversely, stepping into the river causes the agent to 'drown', giving it a highly negative fitness. However, for an agent to individually achieve its goal, it must evolve to perform sub-tasks; a bridge must be built so agents can cross the river safely using Stones, in order to collect their second Resource and thus achieve the goal. To successfully build a bridge, two Stones must be placed in the same Water cell since the river is two cells deep. The aim of the agent is therefore to achieve its goal (i.e. collect both Resources) using the minimum number of Stones; an agent which collects both Resources but places three Stones in the river would thus be less favourable than one which places only two Stones. This bridge-building behaviour is essential if the agent is to succeed; however, developing this behaviour is a complex process as agents must firstly associate the river with a negative fitness, and then that a positive fitness can become accessible *only* if Stones are placed in the river. The challenge of associating specific knowledge or behaviour with specific conditions makes the task difficult for agents to learn to solve, as strong associations must be altered over time; this can potentially lead to the temporary loss of 'safe' behaviour when bridge-building

behaviour is learnt, and therefore can negatively affect the fitness that agents receive while both conditions are learnt.

The RCD testbed is developed in Java, with the configuration of the environment presented in Figure 3.1. Time is measured in 'time-steps', where an agent must move a distance of exactly one cell per time-step in any of the eight cells surrounding the agent's current location. The environment is designed such that multiple agents can exist in the environment, however Figure 3.1 depicts an RCD instance containing two agents; in this case, the agent that starts in the top left of the environment moves first, followed by the agent that starts in the bottom right. In order to explore the effect of interference on agent evolution, one must also observe and compare how agents evolve in an environment alone; in this case, only the agent on the left-hand side of the river exists. Agents are able to interact with all objects within the environment, except from Resource objects which are not allocated to them; non-allocated Resources appear in the environment as obstacles.

Robinson et al. [178] explore how agents evolve to operate in dynamic instances of the original River Crossing task. The focus of this thesis however is on exploring the effect that interference can have on agent evolution and volatility, rather than the ability of agents to navigate dynamic environments; as this has already been demonstrated by Robinson et al. [178], the experimental studies presented in Chapters 4, 5 and 6 instead use a static configuration of the RCD testbed depicted in Figure 3.1. This is not to say that future developments of, or studies using the RCD would be required to also have a static environmental configuration; for the purposes of the studies in the later chapters though, the RCD only uses the characteristics of other instances in the River Crossing family of environments that are necessary for, and directly contribute to, the study of interference.

3.2.1 Gamification of the RCD

The inclusion of two Resources for each agent in the RCD – one immediately accessible, and one initially inaccessible because it is separated by the river – further distinguishes this testbed from others in the River Crossing family. Robinson et al. [178] state that agents in the original River Crossing task are evaluated on three consecutive environments at every generation, where each increases in difficulty; without first showing the agents an environment where a safe passage to achieve the goal already exists, agents were not able to associate the Resource object with a high fitness and evolution resembled random search.

Following this line of thinking, the RCD instead locates one Resource on the side of the river that an agent starts on, in addition to another Resource on the opposite side of the river; if only one Resource was present that was on the opposite side of the river to the agent, there would be no incentive for agents to explore or develop bridge-building behaviour due to the risk of falling into the river.

Furthermore, the RCD environment is designed such that two Stones are required for a bridge to be built; in other River Crossing environments, this is achievable with only one Stone [178, 37, 203]. By increasing the number of Stones to build a bridge to two, the task in the RCD can be gamified such that agents incur an increasing cost for each Stone placed in the river; this cost acts as an incentive for agents to exert the least effort to achieve their goal. However, this also increases the complexity of the task, because agents may accidentally fall into the river when initially experimenting with bridge-building behaviour; agents may consequently endure a period of low fitness while developing the ability to place Stones in the river, making it harder to sustain this behaviour if they do not receive the benefit from collecting the second Resource.

It would be trivial to imagine an environment where there is no cost for placing Stones in the river in the RCD; without this, actions from others within the environment would have little to no impact on agent evolution. Indeed, Robinson et al. [178] show that evolution is able to find solutions in the original River Crossing task much quicker when agents are situated in environments with a one-cell wide river, or where a bridge already exists; solving the task when there is a two-cell wide river however is more difficult. Even with this increased difficulty, if more than one agent existed within the environment, both agents would be able to achieve their goal independently – the actions of others would not influence agent behaviour, evolution, or the fitness they receive. In general, interference from others within the environment is problematic when the fitness agents receive depends on shared environmental features. As the intention behind the RCD is to explore how agents evolve when they may experience interference from others, the cost for placing Stones in the river means that the actions of other agents within the environment *do* in fact affect agent fitness, and the resulting behaviours in which they evolve or maintain. The *social dilemma* that consequently arises when agents share an RCD environment with another agent means that the actions of other agents within the environment – particularly interacting with Stones and thus potentially changing the state of the environment – can affect the fitness that

agents receive. These agents may either complete their task individually and be subjected to the full cost of bridge-building, cooperate to share the cost, or exploit the behaviour of other agents to avoid a cost at all.

Gamification and Agent Fitness

The introduction of an increasing, personal cost for placing Stones into the river to build a bridge creates a Snowdrift Game [152] (also known as the Chicken Game [121, 131, 218, 9] or the Hawk-Dove Game [189, 9]), which is a two-person social dilemma with a cost for cooperation. This means that there is less incentive for agents to cooperate due to the cost of bridge-building, but severe consequences for defection if the agent isn't able to achieve its goal. Gamification adds a subtle complexity to the task incorporated into the RCD environment compared to other testbeds in the River Crossing family of environments, as agents must learn to endure a small cost for a large gain. The fitness, or payoff p for agent i is based on its own individual actions, and is calculated with Equation 3.1, where agents are evaluated on n environments:

$$p_i = \sum_{j=1}^{n} \left(\frac{r_{i,j}}{N_j} - \left[\frac{C_j \times s_{i,j}}{2} \Big(1 + s_{i,j} \Big) \right] - f_{i,j} \right) \tag{3.1}$$

$r_{i,j}$ is the number of Resources collected by agent i in environment j; N_j is the number of Resources allocated to each agent in environment j; C_j is the cost of placing a Stone in the river in environment j; $s_{i,j}$ is the number of Stones placed in the river by agent i in environment j; $f_{i,j} = 1$ if agent i falls in the river in environment j, and is otherwise 0. C and N are constants, with $C = 0.1$ and $N = 2$. Equation 3.1 evaluates each agent's fitness individually – independent of others; this also allows an agent to evolve alone or in a shared environment, as an agent's fitness is calculated solely on its own behaviour. Further, this equation also means that agents can be evaluated on the outcome of one environment ($n = 1$), or multiple ($n > 1$): an agent's fitness is the sum of the fitness achieved in each environment in the evaluation. Chapters 4 and 5 for example explore how agents evolve when evaluated on one environment, whereas Chapter 6 explores the consequences that evolving in many environments can have on evolution and fitness.

Table 3.1 shows a simplified payoff matrix using Equation 3.1, containing commonly observed fitnesses in the RCD. The cost for each Stone placed in the river using this equation

	$S_y = 0$	$S_y = 1$	$S_y = 2$
$S_x = 0$	0.5	0.5	1.0
$S_x = 1$	0.4	0.9	0.9
$S_x = 2$	0.7	0.7	0.7

Table 3.1: Payoff Matrix using Equation 3.1 to show the fitness achieved by agent x in the River Crossing Dilemma testbed, assuming that: agent x has retrieved its Resource object from its own side of the river; agent x will retrieve its second Resource if a bridge has been built; another agent y exists in the environment. S_x and S_y are the number of Stones placed by each agent. $S_y = 0$ also demonstrates the fitnesses that agent x could achieve if it exists in an environment alone.

increases by $C = 0.1$; the total cost of placing one Stone is 0.1, two Stones is 0.3, three Stones is 0.6, etc. The highest payoff when an agent exists in an RCD environment alone is $p_i = 0.7$, as the agent must incur the cost of placing two Stones in the river in order to achieve its goal. In shared environments, agents can receive a payoff of $p_i = 1.0$ by exploiting the other, who receives $p_i = 0.7$ from incurring the total cost of building the bridge on its own. The overall optimal payoff is $p_i = 0.9$ when agents cooperate by sharing the cost of building a bridge. Any fitness below 0.7 indicates the goal is not achieved.

Social Dynamics and the Awareness of Others in the RCD

Cooperation in social dilemmas is influenced by knowing of the existence of a dilemma in the first place [64], and can be negatively influenced if the dilemma's characteristics are unknown or dynamic [218]. Dynamicity however, is an inherent characteristic of coexistence, as the actions of others can change the state of the environment; this makes it difficult to maintain cooperative behaviour as this may become unreliable if the state of the environment (or the behaviour of others) changes. It must be noted that the focus of the studies within this thesis is not on the evolution of social dynamics such as cooperation or competition, but rather how agents can mitigate the effect of interference when socially situated; cooperation or competition can emerge, however this type of behaviour cannot be intended nor understood by the agents, as they are unaware of other agents around them.

3.3 Agent Design

Existing approaches to agent design for the original RCT use a two-tiered neural network architecture, where agents are capable of reacting to dynamic environments (such as a change in environment size or configuration) without needing *a priori* knowledge [178, 35].

As both the state of the agent and environment have the potential to change at each time-step, this two-tiered architecture therefore enables the agent to change its behaviour quickly; this removes the need for planning ahead, as the agent can switch its behaviour as soon as internal or external change is experienced. The agent architecture comprises two neural network tiers, where the first is termed the *deliberative network* and the second the *reactive network*; the combination of these two networks allows agents to make decisions based on their current state, and react immediately to the current state of the environment in line with their current 'decisions' or goals[2].

To keep in line with the design of agents in other studies using the River Crossing family of environments, the studies presented in Chapters 4, 5 and 6 use 'neuroevolution' to evolve the weights of a two-tiered neural network architecture employed by the agents; the agent design in the rest of this section is heavily inspired by the work of Robinson et al. [178] and Borg et al. [37] in particular, which are studies using earlier instances of the River Crossing family of testbeds. Neuroevolution is the process of evolving neural networks using genetic or evolutionary algorithms; this is an alternative to other neural network training techniques such as back-propagation, and is useful for training networks in environments where there is a lack of examples for the network to train from [180, 98]. With neuroevolution, the networks 'train' or 'learn' through the parameters of the network being evolved over time in accordance to some fitness function [202]; the parameters that can be evolved may be purely the weights of the network which are represented as strings of chromosomes [180, 98, 178, 78], or more complex approaches may evolve both the weights and topologies of the networks [240, 200].

It must be noted that neuroevolution using this two-tiered neural network architecture is just one way that agents can 'learn' to solve the task in the RCD testbed; reinforcement learning, learning classifier systems, or even binary strings are just some examples of agent representations that could potentially be explored in the future. Ghouri et al.[3] [96] for example present a *minimal* version of the River Crossing task environment, where the agent representation is a pair of numbers that dictate how many moves the agent will take and in which direction (right or left), rather than the neural network representation used here. Fur-

[2]Note that in other studies using River Crossing testbeds, the entire two-tiered architecture is termed the Shunting Model, and the two aforementioned neural networks are the Decision Network and Shunting Network respectively [178, 37]; the terminology for referring to each tier has been simplified in this thesis to correspond to the behaviours that each layer facilitates (i.e. deliberative or reactive, respectively).

[3]See Footnote 4 in Chapter 2.

thermore, Stanton and Channon [203] expand upon the two-tiered architecture used here, with both a physical and pattern generator network to evolve 3D virtual creatures. Whilst there is not one, single approach to agent design that can be used in the River Crossing family of environments (including the RCD), the research questions that this thesis intends to address are more concerned with *how* agents are able to achieve their goals despite varying environmental factors, rather than focusing on engineering agents to solve this particular problem. As such, this two-tiered neural network architecture is considered sufficient for studying how agents may evolve in the RCD testbed when experiencing interference from other agents. Employing this architecture will also enable the agents to express reactive and deliberative behaviours [178] – a characteristic in which agents designed to operate in most instances of the River Crossing family of testbeds exhibit [178, 37, 203] (the exception to this being the RC- testbed [96], which is a minimal version of the River Crossing task and thus agents have a minimal design to reflect this).

3.3.1 The Deliberative Network

The first tier in this neural network architecture, the deliberative network, generates high-level sub-goals at each time-step based on the current inputs to the network; these inputs correspond to the current state of the agent and the environment. This network is therefore responsible for the decision-making processes of agents; depending on the inputs and the weights of the network, the outputs indicate what the agent decides to do next in terms of sub-goals – attraction to, neutral to, or repulsion from certain objects in the environment. The weights of the network represent the genes or chromosomes of the agent, and dictate the relationship between the state of the agent and environment, and the behaviours in which the agent exhibits. The weights of this neural network are evolved over time with neuroevolution; this process is described in more detail in Section 3.4.

This feed-forward neural network has an input layer with six neurons, a number of hidden layers, and an output layer with three neurons; each neuron in each layer of neurons in the network is connected to each of the neurons in the next layer. Figure 3.2 depicts the deliberative network structure used in Chapter 4, where the number of hidden neurons is inspired by the agent design in the original River Crossing task study [178]. The number of hidden layers and hidden neurons are specified in each study. The weighted sum of the incoming activation to the hidden neurons passes through a hyperbolic tangent activation

function to produce the output. If the incoming signal to the neuron is: within the range [0.35:0.65], the output is 0; less than 0.35, the output is -1; greater than 0.65, the output is 1. Robinson et al. [178], Borg et al. [37] and Stanton and Channon [203] study other instances in the River Crossing family of testbeds, instead using thresholds of -0.3 and 0.3; however, initial experimentation with the River Crossing Dilemma testbed showed that these thresholds inhibited the evolution of agents. If many objects in the environment were 'attractive' at one time, agents had difficulty in navigating towards their sub-goals at each time-step, thus wandering about and being pulled in different directions. By narrowing the range of the thresholds in the RCD compared to previous work, the intention is that agents will have more clearly defined sub-goals at each time-step, depending on their current state and the state of the environment.

The deliberative network's inputs correspond to whether the agent is on Grass, a Resource, Water or a Stone, if it is currently carrying a Stone, and if a bridge has been built partially in the environment (i.e. one Stone in the river out of two). The 'partial bridge' input informs agents anywhere in the environment that a Stone has been placed somewhere in the river; this helps navigation efforts by indicating that some parts of the river are 'shallower' than others, and only require one more Stone to build a bridge. For each of these inputs, the value is 1 if true, or 0 for false. The network then generates sub-goals from these inputs, meaning that the agent can 'deliberate' about what objects it will head towards in the environment based on its state.

The output values of the deliberative network correspond to the sub-goals of the agent, i.e. the resulting behaviour and what the agent will do in the current time-step. The network has three output neurons, each of which correspond to Resource, Stone and Water objects in the environment. After the weighted sum of inputs to the output neurons passes through the hyperbolic tangent function, each of the output neurons will either output a signal of 1, 0 or -1; these values indicate that the agent will be attracted to, neutral to, or repulsed from these objects within the environment.

3.3.2 The Reactive Network

The second tier, termed the reactive network, generates a dynamic activity landscape at each time-step; this activity landscape changes at each time-step depending on the sub-goals generated by the deliberative network, and thus does not evolve like the deliberative

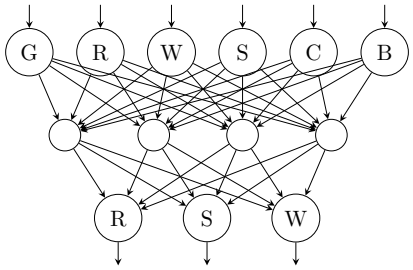


Figure 3.2: The deliberative network is a feed-forward neural network that generates high-level sub-goals. Inputs are 1 or 0, corresponding to the agent's current state: Grass (G), Resource (R), Water (W), Stone (S), Carrying Status (C), if a Bridge partially exists (B). Outputs are 1 for attraction, 0 for neutral or -1 for avoidance for each sub-goal: Resource (R), Stone (S), Water (W).

network. The reactive network is a topologically-organised lattice of neurons with the same dimensions as the environment (in the case of the RCD, this is 19×19), where each neuron is connected to the surrounding eight neurons. Agents can therefore hill-climb towards the goals generated in the previous tier by moving to the cell in its Moore neighbourhood (the surrounding eight cells) with the highest activity at each time-step. Agents must make one move per time-step and cannot remain stationary. Agents also cannot move into a cell occupied by another agent. If more than one of the surrounding cells shares the highest activation, one of these is randomly selected; agent movement therefore has an element of stochasticity, meaning that agents can potentially take different paths when navigating the same environment twice.

Another implementation detail is that an agent will pick up a Stone automatically if it moves onto a cell with a Stone; an agent will also put a Stone in the river automatically if the cell adjacent to it is Water, under the condition that it is carrying a Stone. This is simplified from the process that Robinson et al. [178] employs for example, which uses an additional output neuron in the deliberative network to determine whether the agent will pick up or put down an object. The River Crossing Dilemma testbed however is designed to explore the evolution and behaviour of colocated agents; the additional task of evolving this behaviour of picking up or putting down objects does not contribute to the exploration of the interference that agents experience, and has thus been simplified to focus the studies presented in Chapters 4, 5 and 6.

The sub-goals generated by the deliberative network are used to generate the dynamic activity landscapes in the reactive network. Activity propagates through the reactive neural network using the shunting equation proposed by Yang and Meng [238, 237], which is

characterised by a biologically-inspired equation [100]; this approach was originally used by Robinson et al. [178] to enable agents to express both reactive and deliberative behaviours in dynamic environments. The shunting equation (Equation 3.2) calculates the activity of each neuron in the reactive network at each time-step based on its own activity and the activity of the neurons surrounding it:

$$\frac{dx_i}{dt} = -Ax_i + I_i + \sum_{j=1}^{k} w_{ij}[x_j]^+ \tag{3.2}$$

Alpha A is the passive decay rate, set as $A = 0.2$, which allows the activity of each neuron to decay towards a value of 0; x_i is the current neuron; w_{ij} is the weight of the connection between neurons x_i and x_j, where x_j is one of the surrounding cells in x_i's Moore neighbourhood (indicated by $k = 8$, as each neuron is connected to the surrounding eight neurons); $[x_j]^+$ is calculated by $max(0, x_j)$ – meaning that negative activity cannot propagate through the network. The value of Iota I is dependent on the sub-goals from the deliberative network, and is a large value. For each object in the environment that corresponds to a sub-goal, if the value of the sub-goal is: 1, $I = 15$; -1, $I = -15$; and $I = 0$ otherwise. This creates large hills and valleys in the activity landscape, as inspired by the design of the original RCT testbed [178]. As the reactive network has the same dimensions as the physical environment, the cells in the reactive network which correspond to objects that are deemed to be 'attractive' by the deliberative network are assigned large positive values by this equation, thus creating 'hills'; activity from these peaks then propagates throughout the network, decaying as distance increases from the source of the activation. Similarly, 'repulsive' objects are assigned large negative values to create valleys; activation does not propagate from these, so they will always be avoided. An example of an activation landscape that is generated by the reactive network, using Equation 3.2, is presented in Figure 3.3. This figure shows how agents may 'hill-climb' towards their goals at each time-step, by following a path of highest-activation until they reach their desired goal-object, and avoiding objects which are deemed 'repulsive' by the deliberative network. Note that Equation 3.2 is used exclusively in the reactive network, not the deliberative network.

3.4 Evolutionary Algorithm

Robinson et al. [178] show that agents are able to solve tasks in *dynamic* configurations of the

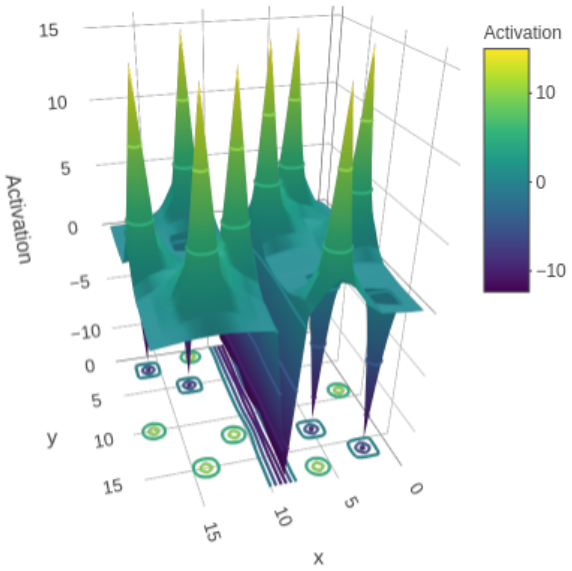


Figure 3.3: The reactive network generates dynamic activity landscapes with Equation 3.2, based on the current sub-goals generated by the deliberative network (Figure 3.2); here, the sub-goals are $[-1, 1, -1]$, meaning the agent is attracted to Stones, and avoids Resources and Water. The activity landscape maps to the physical landscape (Figure 3.1), so agents can hill-climb towards their sub-goals whilst avoiding repulsive objects, by traversing the activity landscape and moving to the adjacent cell with the highest value.

original RCT environment, by expressing both reactive and deliberative behaviours. The agent architecture defined in Section 3.3 is used in the experimental studies in Chapters 4, 5 and 6, and is inspired by that of Robinson et al. [178]; as such, the agents in these studies would be expected to also be capable of solving dynamic configurations of the River Crossing family of environments – the RCD, in particular – as well.

The studies presented in Chapters 4, 5 and 6 evolve agents in the RCD testbed using a Steady State Genetic Algorithm [208], inspired by Robinson et al. [178] and Borg et al. [37]; this means that the population evolves slowly, as one agent is replaced at each generation with the offspring of two parents, rather than replacing large proportions of the population at once. The following common parameters are used in each of the studies in the later chapters. The algorithm evolves a population of 25 randomly initialised agents over a number of generations. At each generation, three agents from the population are randomly selected to compete in a tournament, where they are each given 500 time-steps to achieve their goal in an RCD environment. The evaluation at each generation stops once all agents in the environment reach one of the termination conditions: the agent reaches its maximum amount of time-steps, achieves the goal, or dies by stepping into the river. If an agent achieves its goal, or dies, it makes no further moves until the evaluation terminates.

The worst-performing agent in the tournament at each generation is replaced by an offspring created from the two winners. For each chromosome (layer of weights in the deliberative neural network), this offspring has a probability of $P_{one} = 0.95$ to inherit the chromosome from a random parent (winners of the tournament), otherwise single-point crossover is used. Each connection weight w in the offspring's resulting deliberative network is then mutated by a random value from a Gaussian distribution with $\mu = w$ and $\sigma = 0.01$. Using this algorithm, agents 'learn' on an evolutionary basis, as the genotypes of agents in the population change over time through the recombination and mutation operators defined previously; agents do not learn during their lifetime as their genotype remains unchanged during each generation. However, as there is an element of stochasticity to agent movement (as described in Section 3.3), one genotype can express multiple phenotypes *within* a generation; that is, if an agent is evaluated on multiple RCD instances at each generation, the behaviour expressed by an agent may not be the same in each – even though the genotype is unchanged. In multi-agent environments, only the evolution and goal-achievement of the agent that begins in the top-left corner are analysed; this makes the results from single- and multi-agent environments comparable. The other agent still evolves as described here, however its evolution is not analysed unless otherwise specified.

As the River Crossing Dilemma testbed, the agent design, and the evolutionary algorithms used to evolve said agents in the following chapters are inspired by the original RCT and its extensions, agents are thus evolved for 500,000 generations unless otherwise specified. This is inspired by the experimental setup of the original River Crossing task study by Robinson et al. [178]; 80% of agents evolved to achieve their goal in the more complex environments in approximately 100,000 generations on average, or roughly 450,000 in the worst-case scenario (the simulation was terminated only after 80% of agents achieved their goal). The RC+ task on the other hand, introduced by Borg et al. [37], evolved agents for a maximum of 5,000,000 generations; all agents were able to achieve their goal by building a bridge in a one cell wide river in approximately 500,000 generations on average, whereas in the worst-case scenario the task was completed in around 2,000,000 generations. In both studies, the average number of generations required for the majority of the population to evolve successful solutions increases in line with the difficulty of the environment. In terms of the RCD, 500,000 generations is considered a sufficient length of time for evolution to find goal-achieving solutions, however it would be expected that only a subset of agents

across all runs of an experiment would achieve their goal in this time. This is an adequate length of evolution as the focus of the studies in this thesis is not to find 'the best solution', or to design agents that are excellent at solving this one specific problem; instead, the focus is on analysing and understanding how agents evolve under certain conditions, such that this knowledge can be used in the future to understand and design technical systems that operate in shared environments.

In the remainder of this thesis, the phrase 'agent evolution' is used to describe the process that evolves a *population* of agents. When referencing an individual agent and its evolutionary process, e.g. the evolution of 'Agent A', this in fact describes how the *population* is evolved during evolution, since the population evolves rather than a single agent. This is so that the evolution of two distinct populations of agents can be compared in a more straightforward manner, such as the evolution of 'Agent A' and 'Agent B'.

3.5 Implementing a Testbed to Explore Interference

The approaches listed in Section 2.4.1 are intended to facilitate the development of agent-based systems of varying sizes, through outlining design principles or providing libraries for the creation of custom simulation models. One of the primary aims of the experimental studies in the later chapters of this thesis however, is to study the evolution and behaviour of a small number of agents (i.e. one to study evolution alone, or two to study the effects of interference) in detail, rather than the behaviour and interactions between large numbers of agents. Whilst the latter is still important, it is beyond the scope of this thesis; there is room however to extend the experimental studies to consider large numbers of agents in the future, to observe the effects of interference on a large scale. To address the concerns raised in Section 2.4.1, the experiments are conducted using bespoke testbed and agent implementations – which are one of the contributions of this thesis; this facilitates fine-grained control over the parameters and execution of the experiments, and additionally allows for detailed observation of agent evolution. By designing and creating a bespoke implementation for the experimental studies rather than adopting alternative agent toolkits or design methodologies, any assumptions made by the developers that may potentially affect the design, execution, or results of the simulations, as well as any time required to become familiar with APIs in order to extend existing code, are avoided. The testbed

and the agents created for the experimental studies in this thesis have been developed from scratch using the Java programming language, which is object-oriented and platform-independent. Java was also the programming language of choice for MASON, JADE, and NetLogo for example, because the object-oriented nature of the language is suitable for modelling social phenomena [137] – as is also the case here. This is another reason for using Java to develop the RCD rather than using a predefined approach, framework or toolkit, as these may have limitations or hidden assumptions. Not only this, but all aspects of the testbed and agent design can consequently be controlled, and hence observed when gathering results in greater detail than with any of the approaches discussed previously.

3.6 Analytical Design for Studying Agent Evolution

For each study presented in Chapters 4, 5 and 6, an in-depth analysis is conducted to understand how evolution, goal-achievement, and volatility may be affected by an agent's capabilities or environment. The analytical methods used are described below. The statistical analysis is conducted in each study using the R programming language [173], which is powerful both for statistical analysis and creating visualisations of data. Details about the packages and functions used to conduct this analysis can be found in Appendix A.

3.6.1 Volatility Metrics

Three metrics have been devised to analyse the evolutionary volatility that agents experience, in terms of the fitness of the highest-performing agent in the population over time. Each metric captures different knowledge, and therefore complements the others. The value for each metric is calculated by taking the best-in-population fitness of agents across all runs of an experiment, to capture how much the fitness is prone to change during the course of evolution. These metrics can thus be used to analyse whether a particular approach to agent design, or the effects that evolving in a certain type of environment (such as if the agent exists alone or with another), will cause agent fitness to fluctuate; the more times an agent's fitness fluctuates over time (i.e. higher volatility), the less the agent is able to consistently perform in the environment. This may be because the agent moves between being able to achieve its goals to being unable to achieve its goals, which is not ideal. In this way, low evolutionary volatility would be preferred, as this indicates that agent behaviour –

and consequently the fitness received – is more predictable during evolution. These three metrics are described below.

Standard Deviation over Time (SDoT)

Historical volatility is a common metric used in financial modelling and volatility forecasting that captures the dispersion of values over time, calculated most commonly by the sample standard deviation over a defined time period [170]. This is useful to determine the expected volatility in fitness over time over agents in all runs of an experiment, and as such the SDoT is calculated in the same way: for each run of an experiment, the standard deviation is calculated from the best-in-population fitness at each generation over the course of evolution.

Cumulative Absolute Change over Time (CACoT)

To quantify how an agent's fitness changes over the course of evolution – and by how much – the CACoT metric is introduced to capture the magnitude of the changes that the best-in-population fitness of an agent endures over time. Here, the count is incremented by the absolute change in fitness between generation g_{i-1} and g_i. A high CACoT therefore indicates that fitness fluctuates by large amounts; often, this will indicate that an agent tends to alternate between receiving high fitnesses when achieving its goal, and low fitnesses where it either cannot achieve its goal or fails the task (in terms of the RCD, this would be by the agent 'drowning' by stepping into Water).

Count of Change over Time (CCoT)

Complementing the CACoT metric, the *number* of times that the best-in-population fitness over evolution can be captured with the CCoT metric; here, the metric increments by one only if the fitness in generation g_i is not equal to g_{i-1}. The CCoT metric therefore captures purely the number of times the fitness of an agent fluctuates during evolution, rather than the magnitude of those changes. A high CCoT indicates that the agent experiences many fluctuations in fitness during evolution, implying that the agent cannot maintain its fitness.

3.6.2 Statistical Analysis

In addition to the volatility metrics described above, statistical analysis is conducted for the results of each experiment to understand the evolution of agents in each experiment

further. This analysis is conducted for the fitness agents receive during evolution, as well as the volatility agents experience, using the metrics defined in the previous section.

Statistical Moments

In statistics, the shape of a distribution of a variable can be described using four moments: mean, variance, skewness, and kurtosis [87]. These moments are calculated across all runs of an experiment using the R package `moments` [123], to quantify the distribution of the values of either the fitnesses achieved by agents in each run, or the volatility captured by each of the three metrics defined in the previous section. The moments of the distributions of two experiments (for example, when an agent evolves alone compared to when it evolves in a shared environment) can thus be used to compare the results of each experiment, and to indicate whether one approach may be preferable over another.

The mean captures the expected value of a distribution, whereas the variance describes the dispersion of values around the mean. Skewness and kurtosis are used to describe the shape of the distribution further, giving an indication of how much the distribution varies from the normal distribution [87].

Positive skewness (or right-skew) indicates that the peak of the distribution is towards the lower end of the scale, with a longer tail towards the higher end of the scale. Negative skewness (or left-skew) indicates the opposite, where the majority of values are towards the higher end of the scale, with a tail extending towards the lower end of the scale.

Kurtosis however is a measure of the tailedness of a distribution, and therefore how likely outliers are. A normal distribution has a kurtosis of 3 [65]; the kurtosis of other distributions is described in comparison to the kurtosis of a normal distribution. A 'mesokurtic' distribution has a kurtosis of 3, so the kurtosis is the same as that of a normal distribution. 'Leptokurtic' distributions have positive kurtosis, meaning that the value of kurtosis is greater than 3. 'Platykurtic' distributions have negative kurtosis, as the distribution has lower kurtosis than a normal distribution, and thus kurtosis is less than 3. Positive kurtosis indicates that the distribution would have more outliers and extreme values than distributions with negative kurtosis.

In addition to these four statistical moments, the median of each distribution is captured as this is useful to compare the central values of the distributions in conjunction with the mean. Further, observing both the mean and median can often indicate that a distribution

is skewed; a higher mean than median can indicate positive skewness, whilst the opposite indicates negative skewness.

Statistical Tests

Whilst the statistical moments described in the previous section can be used to describe the distributions of the fitness and volatility metrics in each experiment, statistical tests can be used in addition in order to make concrete statements when comparing one distribution to another. The statistical tests outlined below are conducted using the appropriate functions supplied by the `stats` package, which is part of the core R distribution [173].

Many statistical tests exist which compare different characteristics of the distribution, or make different assumptions about the distribution. To determine which statistical test to use, a Shapiro-Wilk test for normality is first conducted for each experiment; this is powerful for a wide range of distributions [241], including those that are symmetric or asymmetric, or with high or low kurtosis. For this reason, this test is deemed suitable over other normality and goodness-of-fit tests such as the Cramr-von-Mises and chi-squared tests, as these are less powerful for such a diverse range of distributions. For the Shapiro-Wilk normality test, the null hypothesis is that the data is normally distributed; if the result of the test is $p < 0.05$, there is evidence to reject the null hypothesis that the distribution is normal.

After considering whether the distribution is normal or non-normal, an appropriate statistical test can be chosen. In Chapters 4, 5 and 6, Wilcoxon Signed Rank statistical tests are conducted, as these non-parametric tests do not assume that the data is normal [232, 87]. This test compares the medians of two distributions, where the data is 'paired'; this means that the distributions of related samples can be compared. Specifically in this thesis, agents evolve with a 'standard' evolutionary approach, or with a sociologically- or biologically-inspired approach; these different approaches can thus be compared directly, as the same populations of agents evolve with different approaches. For each experiment, one two-tailed and two one-tailed Wilcoxon Signed Rank tests are conducted; the null hypotheses for these tests are that the distribution medians are equal, and that there is no directional difference in the distribution medians respectively. These null hypotheses can be rejected if the p-value obtained from the test is below 0.05, which indicates that the result is statistically significant. If the data were found to be normal when conducting a Shapiro-Wilk test, a parametric alternative such as the Student's t-test would instead be considered, which

compares the mean of two distributions.

Effect Size Estimates

Statistical tests can be used to ascertain whether two approaches are distinct from one another, but these tests do not shed light on *how different* the approaches are to one another. Effect size estimates can thus be used in addition to both the statistical moments and tests described above, in order to quantify the relationship between two variables, and therefore the magnitude of the effect between the two approaches. The correlation coefficient r [214] can be used to estimate such a magnitude between two variables; this effect size r can be estimated using the following formula, as defined by Rosenthal [181]:

$$r = \frac{Z}{\sqrt{N}} \tag{3.3}$$

where Z is the z-score (also known as the standard score, or z statistic), and N is '*the number of subjects or other sampling units*' [181] (for example, if an experiment is repeated 100 times, then $N = 100$). This particular method for estimating the effect size is chosen because it is suitable to use in conjunction with the non-parametric Wilcoxon Signed Rank statistical test. The `wilcoxonZ` function provided by the R package `rcompanion` [142] is used to calculate the z-score for a Wilcoxon Signed Rank statistical test, which can then be used in Equation 3.3 to calculate the effect size. Other effect size estimates such as Cohen's d [58] exist, however Rosenthal [181] notes that r is more versatile in its usage.

Cohen [58] suggested three different categories that can be used to describe the strength of the relationship between two variables, captured by the effect size estimate r (Equation 3.3): $r \geq 0.1$ is small, $r \geq 0.3$ is medium, and $r \geq 0.5$ is large. The effect size r ranges between -1.0 and 1.0, where $-r$ indicates a negative relationship between the two variables, and $+r$ indicates a positive relationship. For example, if the effect that agent design B has on fitness is compared to the 'standard' agent design A, an effect size of $r = -0.5$ indicates that agent design B has a large effect on agent fitness; as r is negative, agent design B increases fitness. The effect size estimates between different experiments can thus be compared, as the strength of the relationship between the two variables in each experiment can be contrasted when the agents in the studies are subjected to different environmental conditions.

3.7 Conclusion

The River Crossing Dilemma (RCD) testbed presented in this chapter has been designed to address the shortcomings of the River Crossing family of environments, in that the RCD contributes the ability to explore how multiple agents evolve to achieve their individual goals. By incorporating a social dilemma into the RCD, the actions that agents take within the environment can have an impact on the other agents, as the state of the environment changes. Consequently, the effect that interference can have on agents can be studied in close detail, and the evolutionary volatility that agents experience as a result of this interference can also be quantified.

The contributions of this chapter are:

- A gamified testbed – the River Crossing Dilemma – specifically designed to observe how interference affects agent evolution and goal-achievement, where cooperation and exploitation can emerge but cannot be intended.
- An in-depth description of the testbed, agent and evolutionary algorithm design.
- A comprehensive approach to analysing the results obtained from the experiments conducted in the River Crossing Dilemma testbed.
- Three metrics that measure the evolutionary volatility that agents experience, based on how frequently or how much an agent's fitness fluctuates during evolution.

A core characteristic of the River Crossing family of environments – and indeed the River Crossing Dilemma – is that the testbeds are extensible, and may be made arbitrarily complex. As a result, the RCD is versatile in the sense that agents can be observed in both single- and multi-agent environments, such that the effect of evolving in either a shared or individual environment can be contrasted. As agents have no capacity to perceive or learn of one another, the RCD simulates the conditions in which components of sociotechnical systems in the real-world experience when sharing environments with potentially unknown other systems. Experimentation within the RCD testbed should aid understanding of the implications of interference, and the consequences that it can have on goal-achievement and performance. Consequently, *how* systems, and the artificial agents that simulate them, may combat this interference can be investigated; the experimental studies presented in

Chapters 4, 5 and 6 therefore use the RCD testbed to observe the impact that sharing an environment can have on individuals, as well as exploring how nature-inspired approaches to agent design may affect agent fitness or performance, and the resulting evolutionary volatility experienced.

Chapter 4

Traditional Action and Evolutionary Volatility

The work presented in this chapter has been adapted from the following publications:

[15] C. M. Barnes, A. Ekárt, and P. R. Lewis. Social Action in Socially Situated Agents. In *Proceedings of the IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 97–106. IEEE, 2019. doi: https://doi.org/10.1109/SASO.2019.00021.

[18] C. M. Barnes, A. Ekárt, and P. R. Lewis. Beyond Goal-Rationality: Traditional Action Can Reduce Volatility in Socially Situated Agents. *Future Generation Computer Systems*, 113:579–596, 2020. doi: https://doi.org/10.1016/j.future.2020.07.033.

4.1 Motivation

As society delegates more decisions to intelligent machines, and interactions between such machines that exist in shared environments becomes more prevalent, capturing elements of evolved human social behaviour will be increasingly important. Bellman et al. [28] state that systems require social awareness to *intentionally* cooperate, coordinate, or act socially, and perceive and reason about others; however, social action and social self-awareness are two essential aspects that so far remain largely unexplored. Human societies have no global knowledge or central point of control; how, then, do humans interact effectively? Organic Computing approaches this by observing and controlling a group of interacting, self-organising entities [155]; in this chapter however, a microsociological approach is explored as a step towards socially intelligent systems capable of social awareness at the individual

level. Humans have evolved the ability to achieve goals in complex social systems by considering others and acting socially; inspired by this, the experimental study conducted in this chapter operationalises social action theory [227] in socially situated agents that pursue individual goals in shared environments. To concentrate the scope of this study, two types of social action are operationalised: 'traditional action', which is acting similarly to the rest of the population; and 'goal-rational action', which is seen in current systems where the most effective action to achieve a goal is taken. Complementing this, 'random action' is introduced, which adds randomness into the behaviour of agents by way of Random Immigrants [57]. The aim here is to distinguish the differences between goal-rational and traditional action, and additionally investigate whether these approaches are quantitatively different to introducing Random Immigrants into the population [57] – a well-established mechanism widely used in dynamic optimisation problems and in dynamic environments to add diversity to populations in genetic algorithms [190, 236, 244, 130, 93]. The need to act in a socially-sensitive way challenges the assumption that goal-rationality is necessary for agents to achieve goals in shared environments with less than complete knowledge.

The experiments are conducted using the River Crossing Dilemma testbed (Chapter 3), which was designed to explore arbitrarily complex problems in shared environments. Firstly, the effect that interference can have on agents that are able to achieve individual goals alone is explored, to assess how learnt knowledge is maintained. These results are then generalised over many experiments for agents that begin evolution with no prior knowledge. Conclusions are finally drawn to compare the volatility of evolution with each type of social action, and whether goal-achieving behaviour can be learnt and maintained despite interference.

The remainder of this chapter is organised as follows: Section 4.2 explores the theory of social action, and contextualises this in terms of computational systems; the design of the experimental study is then presented in Section 4.3, which details how social action theory is operationalised in agents that evolve in the RCD; the results of said study are presented in Section 4.4, exploring how coexistence can affect evolution, and the impact that social action can have on these socially situated agents; analysis into how agents in the study evolve is conducted in Section 4.5, which also delves into whether a trade-off between fitness and volatility exists between the various types of social action employed in the study; finally, Section 4.7 concludes the chapter by discussing the findings and summarises the contributions.

4.2 Acting Socially in Shared Environments

Following on from the discussion in Chapter 2, interference means that actions can have unintended or unanticipated consequences on both the actor and others in shared, complex and dynamic environments [147], making shared and individual goals more difficult to achieve. Humans overcome these issues by acting *socially* and not purely *individualistically* in social environments. Many computer science researchers have therefore been inspired by theories of psychology, sociology and cognitive science, such as in organic computing [155], self-awareness [133, 125], and social dilemmas, social learning, altruism and agent societies [109, 116, 226, 164]. It thus seems logical to draw parallels between the exploration of human social phenomena in sociology and socially situated agents; taking inspiration from how humans act in society may provide a similar benefit to artificial agents that operate in shared, and thus dynamic, environments.

If an agent cannot perceive or learn about potentially unknown others, or the effect that their actions can have on itself, interference can affect how goals are achieved in ways that cannot be understood; without a mechanism to overcome interference and unanticipated events beyond their control, agents will be unable to make appropriate decisions at runtime in accordance with their goals. It is therefore necessary, in broader terms, to move towards socially situated technical systems with the capacity for social awareness [28], by equipping them first with the ability to maintain goal-achieving behaviour despite unanticipated interference from other systems. As humans have evolved to both exist and thrive in diverse environments shared with many others through acting socially, this study operationalises the theory of social action, proposed by Weber [227], to understand how agents may also benefit by acting socially when coexisting with others. As highlighted by Bellman et al. [28] and Castelfranchi [51], designing systems that are capable of acting in a socially intelligent manner is becoming increasingly necessary as modern computer systems are evermore situated in shared environments. Operationalising social action theory is therefore one way to explore how artificial systems may operate in a human- or socially-inspired manner, in order to mitigate the effects of interference from others. Endowing systems with the ability to act in a more socially-oriented way, regardless of whether they exist in isolation or not, may help them to perform consistently *without* requiring extensive knowledge of others.

4.2.1 Defining Situatedness

To explore how artificial agents may behave in shared environments, one must first adopt and define relevant terminology to describe the concepts that arise as a result of coexistence. Rao et al. [175] define 'situated agents' as those that are resource-bound, operate within and continuously interact with dynamic environments, and balance reactivity with deliberation. Lindblom and Ziemke [136] use the term 'social situatedness' to refer to the concept that the behavioural and cognitive processes of agents are affected by both the social and cultural aspects of the environment in which they are situated; further, an agent that is 'situated' is said to be coupled with its environment, in that the agent can affect the environment in some way and vice versa. This is influenced by the theories of Vygotsky [224], such that the social aspects of environments that agents are situated in affect the individual intelligence of the agents. Leading on from this, Dautenhahn et al. [63] describe 'socially situated agents' to be those that not only gather information from their physical environment, but from the social component of it as well.

The intention behind the study presented in this chapter is to investigate the impact that evolving in a shared environment, as opposed to an environment that is inhabited by a single agent, has on the ability to evolve, and how agents are able to pursue individual goals. As agents may potentially share an environment with another, the term *socially situated* is adopted to describe said agents that evolve and coexist with another, unknown agent within the environment. Due to their operation in a dynamic and shared environment, an agent's behaviour can interfere [51] with the actions and goals of others, where the knowledge acquired is influenced by the physical and social environment – whether the agents are aware of it or not. If an agent cannot perceive the cause of environmental changes to be the result of the actions of another agent, then the changes will be perceived as uncertainty arising from the environment itself; agents in this situation will still be 'socially situated', as the way they evolve will be influenced by the consequences of the actions of others regardless of whether they are aware of the existence of others.

4.2.2 Defining Social Action

By establishing the concept of 'socially situated agents' and understanding that coexisting agents may interfere with one another, it becomes easier to see similarities between how

artificial agents may evolve in shared environments and how humans have evolved to behave in society. Weber [227] theorised that humans are capable of *social action*, which are actions oriented towards, and that consider the behaviour of, others (this is described in more detail in Section 4.2.3); these actions hold a 'subjectively understandable' meaning. 'Meaning' in this context refers to the motivations of the actor, such that the motivations can be understood from the actor's perspective: *why* an actor acts in a particular way, given its own perspective and circumstances. By extension, artificial agents could also be capable of social action following this definition, if sharing an environment with another agent.

Weber [227] goes further to define what does and does not constitute a 'social action': actions with inanimate objects do not involve other actors, and are thus not social; actions that have no meaning, i.e. those without motivation or deliberation, are not actions, but merely behaviours. To further distinguish between these concepts, Sztompka [209] proposes a hierarchy of social action, a subset of which is outlined below:

- *Behaviour* is automatic, reactive and reflexive.
- *Action* is intentional and purposive, with meaning to the actor.
- *Social Behaviour* holds no meaning to the actor. It is reactive, therefore no deliberation occurs. A behaviour becomes *social* when directed or oriented towards another.
- *Social Action* holds meaning to the actor, and is intentional. A rational decision is made to act in a certain way, taking into account different factors such as the actor's emotional state and the current situation. An action becomes *social* when directed or oriented towards another.
- *Social Interaction* requires a response to a social action from another actor.

In this chapter, the experimental study defined in Section 4.3 considers how agents that have no capacity to perceive or learn of other agents in their environment may evolve to pursue individual goals, using the River Crossing Dilemma as a testbed. Using this terminology as outlined by Sztompka [209], these agents would not be capable of social interaction, as it is assumed that a response from one actor to another would require awareness of the other – as well as the action itself. The distinction made between behaviours and actions is the *intent*; behaviours are reactive whereas actions are deliberative or 'hold meaning'. The

other way these concepts are categorised in this hierarchy is with orientation: an action or behaviour becomes 'social' if it is directed, or oriented towards another. The agents in this chapter would therefore be capable of social action if an action is intended, and is oriented towards another; Section 4.3 elaborates on how social action is specifically operationalised.

4.2.3 Ideal Types of Social Action

By the definitions proposed by Sztompka [209], a 'social action' is one that is intended or deliberate, and is oriented towards another. Preceding this, Weber [227] outlined his theory of social action, which defines four *'idealtypus'* of social action; these ideal types describe the motivations behind social actions in a simplified model, to aid analysis of complex human actions. The word 'ideal' in this context is not synonymous with 'perfection', but an 'idea', in order to conceptualise the different types of social action [227]. An 'ideal type' does not exist in isolation in practice, and is a simplification of a real-world concept so it can be theorised about. Weber's four ideal types of social action are described below.

Instrumental-Rational Actions are chosen for their effectiveness in achieving a *goal*, and are justifiable from the perspective of the actor; consequently, these are often termed *goal-rational* actions. Other goals, the range of possible actions, and the consequences of performing the action are considered to decide the most appropriate action; the meaning is tied to the end result, as the action is chosen with the goal in mind. Most artificial agents, especially in machine learning, are instrumental-rational by this definition.

Value-Rational Actions are determined by the values or *beliefs* held by an actor, such that performing the action itself carries meaning instead of the outcome. Actions are rationalised in terms of *ethical* or *religious* beliefs, or to any cause valued by the actor. Rationality is a justifiable, conscious decision of how to act, and is understandable when considering the motivations of the actor; actions may seem irrational to outside observers if the motivation is not immediately clear. Pure value-rational action is where the value outweighs the consequence of the action, such as a soldier sacrificing themselves for another.

Affective Actions are reactive and impulsive actions in response to an *emotional* state or exceptional stimulus. Affective action appears inherently irrational, as the consequences of the action may not be considered and thus may be difficult to justify. An example is striking someone out of rage.

Traditional Actions are *habitual*, or those in reference to a cultural *custom*; it can therefore be seen as acting in the same way as others. These can be described as mindless, automatic, or ritualistic actions; there is no obligation to act in this way, rather the rationalisation for performing the action is that 'it has always been done this way'. An example is using specific eating utensils; deliberation reduces over time as the action becomes second nature. Traditional actions therefore might not necessarily be optimal or most appropriate in terms of achieving a goal, but can be adequate means to satisfy the end.

4.2.4 Social Action in Computational Systems

Whilst not explicitly used to define current approaches, the theory of social action [227] can be applied to computational systems with actions that are determined by error-function-based learning or objective-function-based search for example; these would be considered 'goal-rational', as they are engineered to maximise their ability to achieve a particular goal. However, as this chapter will demonstrate through an experimental study, there are unintended consequences associated with the actions of goal-rational agents that coexist in a shared world; this can manifest as volatility in evolution and a loss of ability to achieve one's own goals. Human evolution has favoured social behaviour to deal with issues arising from living in the presence of others [192]; without this, humans struggle to adapt or survive [62]. Weber's social action theory [227] has thus been adopted in this chapter to describe how agents evolve in shared environments; consequently, this goal-rationality seen in current systems can be compared to other, less commonly observed types of social action, to explore how computational systems may begin to overcome these issues in a human-inspired way.

Social action can be operationalised in different ways. Value-rational action would become especially critical when systems make decisions on behalf of humans, and would differ between systems depending on the values of the environment it is situated in; this poses the question of how one can trust that the decisions made align with human values. Affective action could be taken when one does not know how to proceed in an unknown situation; as such, all actions may appear to be irrational unless abstracting previous knowledge to justify the decision of performing the action. Traditional action could simply be copying or imitating what most others are doing, forming traditions over time.

This chapter explores the notion of traditional social action and its effect on goal-achievement in socially situated agents, inspired by Weber [227]. It is operationalised as an

action derived from a representative state of the population at specific points in time. To align with the definition of 'social action' above, a social action here is oriented towards the others in the population in terms of the evolutionary process, and not the other agents that may exist within the environment. As agents in this study cannot perceive the presence, actions or intentions of other agents, traditional action intentionally oriented towards others in the environment would be a topic of future explanation in agents that *are* able to perceive and reason about others. This chapter provides an in-depth analysis of how goal-rational and traditional action can affect agent evolution; value-rational and affective action are thus beyond the scope of this chapter. However, Bellman et al. [28] note that capturing human values in computational systems without degrading the value itself is a challenge that needs to be addressed to design socially aware systems, making the concept of value-rational action both an exciting and important area to study outside the scope of this thesis.

In addition to traditional and goal-rational action, 'random social action' is also introduced, which adds Random Immigrants to the population [57]; adding Random Immigrants is an established approach to increase population diversity in evolutionary algorithms. To this end, a comparison can be made between how diversity affects agent evolution, and whether social action within the evolutionary process can mitigate the effect of interference without an explicit awareness of what is causing it (i.e. other agents in the environment), as a step towards socially aware agents.

4.3 Experimental Study

The experimental study in this chapter is designed to explore how agents evolve in environments either on their own, or shared with another, unknown agent. This is to ascertain the extent to which interference from the actions of other agents affects how agents evolve, and the resulting evolutionary volatility experienced by agents. Agents are evolved using the evolutionary algorithm detailed in Section 3.4 in the River Crossing Dilemma (RCD) testbed as described in Chapter 3. To challenge the goal-rationality that is seen in most current systems, traditional action is operationalised to observe whether social action has any benefit on agent evolution or volatility. Further, random action is used to compare the two types of social action – goal-rational, and traditional action – to an established method to increase diversity in evolutionary algorithms.

4.3.1 Agent Design

The agents in this study are designed using the two-tiered neural network architecture described in Section 3.3. Specifically for this study, the deliberative network has one hidden layer of four neurons that connects the six-neuron input layer and the three-neuron output layer; this can be seen in Figure 3.2. All other details are as previously specified.

4.3.2 Operationalising Social Action in the Evolutionary Algorithm

The notion of social action is operationalised within the evolutionary algorithm, where a social action is oriented towards others in the population in terms of evolution, rather than other agents that may exist in the environment. Specifically, the offspring produced at each generation is dependent on the type of action used.

In this sense, the evolutionary algorithm defined in Section 3.4 is *goal-rational*; by creating an offspring from the winners of a tournament, the offspring is likely to contain genetic material that enables it to achieve a high fitness like its parents, thus maximising the algorithm's potential to evolve goal-achieving individuals. This description of 'goal-rationality' can indeed be extended to evolutionary algorithms in the general sense: they evolve populations of individuals towards better areas of the search space over time, using recombination, mutation and selection operators, where higher-fitness individuals are usually given more opportunity to reproduce than lower-fitness individuals [12].

To ascertain whether other types of social action may be more beneficial to agents that experience interference, traditional action is further operationalised as follows; at each generation, there is a 90% chance for the worst-performer of each tournament to be replaced by the current goal-rational offspring of the best two parents (as outlined in Section 3.4), and a 10% chance for it to be replaced with an offspring that is a representative state of the population. The genetic material for this traditional offspring is captured by calculating the median for each weight in the deliberative network across all agents in the population. Since the agents use a two-tiered neural network architecture to make decisions about how to move in the RCD environment, two agents each with different weights in their deliberative network may act in the same way; in other words, many genotypes can produce the same or similar phenotypes. Whilst there are many potential ways to operationalise tradition, this particular method was chosen as a simple means of establishing traditions of phenotypic

behaviour across multiple genotypes, that can potentially change during evolution. This 'traditional action' is thus a blend of both traditional and goal-rational action.

Random action, which is introduced to compare both traditional and goal-rational action to an established method of increasing population diversity (Section 4.2.4), is operationalised similarly to traditional action. At each generation, there is a 10% chance for the worst-performer of the tournament to be replaced with an agent with a randomly-initialised set of weights, and a 90% chance for it to be replaced with the standard goal-rational offspring as defined in Section 3.4. This small chance of replacing the worst agent with a random solution would increase population diversity, where this random solution is formally known in the literature as a 'random immigrant' [57]; this would allow the algorithm to escape local optima, and to traverse unexplored regions of the search space. This 'random action' is a blend of both random and goal-rational action.

Preliminary investigation of varying replacement probabilities showed that little effect or benefit was seen with a replacement probability lower than 10%; forming and maintaining traditions with traditional action is more difficult with lower replacement probabilities, as solutions have little time to influence the population. Further, higher probabilities such as 20% or 30% showed that the population became saturated and diversity was reduced, meaning that it was increasingly difficult for evolution to explore the fitness landscape. A 10% replacement probability was also deemed suitable because in the case of random action, higher probabilities of replacement start to resemble random search, rather than evolutionary search.

4.3.3 Experimental Design

Agents are evolved using the Steady State Genetic Algorithm specified in Section 3.4, for either 500,000 generations when alone, or 1,500,000 generations when sharing an environment. As outlined in Section 3.4, the average number of generations it takes for evolution to find goal-achieving solutions in other River Crossing studies [178, 37] increases with the difficulty of the environment. As it is hypothesised that evolving in a shared environment would be more difficult than when alone because of interference, the length of evolution is thus increased. Further, the effects of interference can be observed over a longer period of time as a consequence, which will shed light on how agents evolve when their behaviour and evolution is influenced by the actions of others in which they evolve alongside. Where

agents are evolved in a shared environment, two separate populations (one for each agent in the shared environment) are evolved alongside one another.

The study presented in this chapter is split into two parts: firstly, individual agents are observed in the RCD to understand how interference may affect evolution in detail; secondly, a broader approach is taken to understand how interference affects agents in general, by analysing the way many agents evolve, and how the type of social action implemented can affect evolution. A summary of the experiments is presented in Table 4.1.

Part One: Exploring the Effect of Interference

The first set of experiments, presented in Section 4.4.1, explore whether ten randomly-initialised agents are able to achieve individual goals with goal-rational action; these evolved alone for 500,000 generations.

The effect of interference is then explored in a further two sets of experiments, which observe whether agents are able to continue to achieve goals either with or without continued evolution in a shared environment. Firstly in Section 4.4.2, the ten evolved agents from the first set of experiments are randomly arranged into ten pairs to observe the effects of interference in a shared environment; these agents initially evolve alone for 500,000 generations (Section 4.4.1), and then continue to act in a shared environment for a further 500,000 without further evolution. As agent movement is stochastic, these agents have the potential to act differently in each generation despite their genotype remaining the same; it would therefore be expected that agent fitness over these 500,000 generations would not be static, but would in fact often fluctuate since agents may act differently in each generation. Secondly, Section 4.4.3 then arranges the same ten agents from the first set of experiments into 30 random pairs; these pairs are evolved together in shared environments for 1,500,000 generations after their initial period of evolving alone, to observe whether goal-achieving behaviour can be maintained despite interference. These experiments are thus labelled 'continued evolution'.

Section 4.4.5 then compares goal-rational action (Section 4.4.3) with the introduction of traditional action in the same 30 pairs of agents; this is to ascertain whether traditional action may help to preserve goal-achieving behaviour when agents that are able to achieve their goals when alone experience interference for the first time. These agents are evolved for 1,500,000 generations after their initial period of evolving alone.

	Section	Action Type	Evolution Type	Agents	Generations	Runs
Part 1	4.4.1	Goal-Rational	Evolving Alone	1	500,000	10
	4.4.2	Goal-Rational	No Evolution	2	500,000	10
	4.4.3	Goal-Rational	Continued Evolution	2	1,500,000	30
	4.4.5	Goal-Rational & Traditional	Continued Evolution	2	1,500,000	30
Part 2	4.5.1	Goal-Rational	Evolving Alone	1 (1G)	500,000	100
	4.5.1	Goal-Rational	Evolving Together	2 (2G)	1,500,000	100
	4.5.1	Goal-Rational & Traditional	Evolving Alone	1 (1GT)	500,000	100
	4.5.1	Goal-Rational & Traditional	Evolving Together	2 (2GT)	1,500,000	100
	4.5.2	Goal-Rational & Random	Evolving Alone	1 (1GR)	500,000	100
	4.5.2	Goal-Rational & Random	Evolving Together	2 (2GR)	1,500,000	100

Table 4.1: Experiment breakdown, outlining the section number the experiments are presented in, the type of action used, the type of evolution, the number of agents in the environment, the number of generations, and the number of times the experiment is repeated.

Part Two: Exploring Social Action

The second part of the study is a more coarse-grained investigation of how agents evolve with social action, and how interference affects agent evolution in general; this is compared to the more fine-grained approach in the first part of the study, which looks at how *individuals* evolve, rather than the general characteristics that arise in the evolution of many agents.

The effect that traditional action has on agent evolution is explored on a broader scale in Section 4.5.1; 100 goal-rational agents, initialised with random weights, are evolved with and without traditional action in both individual and social environments. To ensure the results in shared environments are not biased against a particular agent, agents are assigned a randomly-generated partner. The agents are evolved for 500,000 generations if they are alone, and 1,500,000 generations if socially situated.

Section 4.5.2 finally ascertains whether traditional action is quantitatively different to the introduction of Random Immigrants [57]. The same 100 agents as in the previous set of experiments are evolved with random action instead, in both individual and social environments. Finally, an analysis is conducted to ascertain whether a trade-off between fitness and volatility exists for the three types of social action.

4.4 Results

4.4.1 Evolving Alone with Goal-Rationality

Agents do not experience interference when they are in an environment alone. As such, they can achieve individual goals with independent asocial learning with goal-rational action.

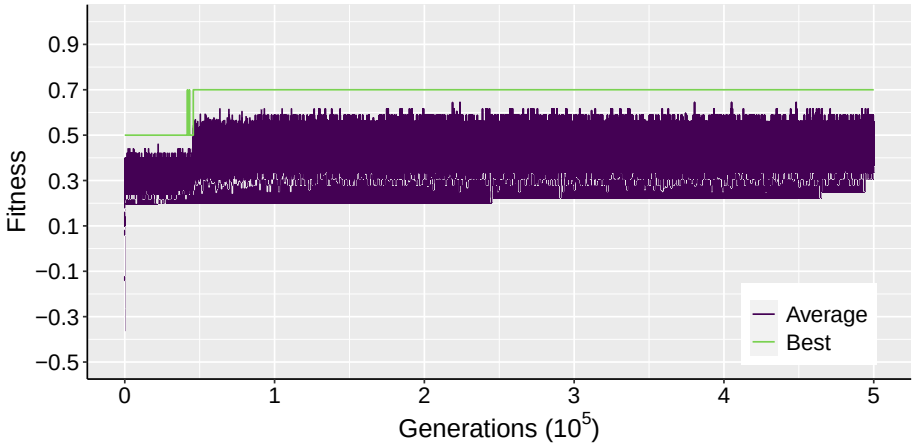


Figure 4.1: Agent F can achieve its goals when alone in an environment; it initially learns to collect one Resource to get a fitness of 0.5, then to build a bridge to achieve its goal around generation 50,000, giving it a fitness of 0.7.

Figure 4.1 depicts the fitness of a single agent during evolution; this agent evolves goal-achieving behaviour at around generation 50,000, which is then maintained throughout the rest of evolution. This agent has thus evolved to build a bridge in the river with two Stones, which enables it to cross the river and fully achieve its goal. Using Equation 3.1, the best-in-population fitness of 0.5 at the beginning of evolution indicates that the agent retrieves one Resource from its own side of the river, and does not endure any cost from placing Stones in the river. Once goal-achieving behaviour is evolved, the fitness increases to 0.7, which indicates that agents retrieve both Resources in the environment, giving a fitness of $(2 \times 0.5) = 1.0$, while the cost of building the bridge is deducted.

Following this, the evolution of ten goal-achieving agents is explored, where the average population fitness of these ten agents is presented in Figure 4.2. Whilst not impossible to achieve, this task initially appears difficult to solve simply because the fitness function does not 'lead' agents towards their goals with incremental rewards; agents encounter a very large, neutral network landscape during evolution as a result. In each experiment, goal-achieving behaviours were maintained once learnt, and in all ten agents, the goal was achieved and maintained by generation 50,000; this can be seen by the increase in mean fitness in Figure 4.2, which rises and then is maintained at this point during evolution. Random mutations during the breeding process periodically create agents with lower fitnesses than the best agent in the population. These lower-fitness offspring may fall in the river for example, thus reducing the fitness average; however, these solutions are replaced

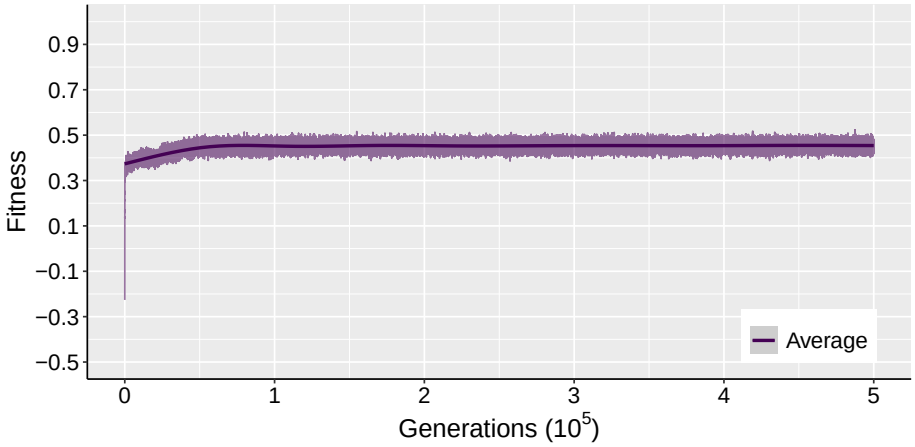


Figure 4.2: The mean population fitness of ten agents evolving alone. All ten agents that evolved alone sustained the behaviours necessary to achieve their goal by generation 50,000.

quickly, leaving the beneficial behaviours to remain. These ten, individually evolved agents are henceforth labelled Agents A through J. These experiments demonstrate that agents are able to evolve and maintain goal-achieving behaviour when they evolve alone with goal-rational action; these results can therefore be compared to agents that evolve in shared environments, to observe the effect of interference on evolution.

4.4.2 Coexistence without Continued Evolution

The ten agents in the previous section are then arranged into pairs, to observe how goal-rational action affects agents in shared environments both without (Section 4.4.2) and with (Section 4.4.3) continued evolution. Table 3.1 shows common fitnesses and their associated behaviours: cooperation (0.9), exploitation (1.0), and achieving the goal individually (0.7).

In this set of experiments, agents that evolved for 500,000 generations alone in the previous section are then placed into shared environments *without* further evolution; simply, the genetic code of each agent remains unchanged, such that the effect of interference can be explored in agents that have already evolved the ability to achieve their goals. Three emergent dynamics are observed in these non-evolutionary experiments: one agent exploits the other for a higher payoff (Figure 4.3a); both agents co-exist and achieve their goals similarly to when they are alone (Figure 4.3b); or one or both agents cannot achieve their goals (Figure 4.3c). In the latter case, the interference experienced, arising from the actions of the other agent in the environment, may cause agents to continue putting more Stones in the river which accrues a larger cost; it may even alarmingly make them walk into the

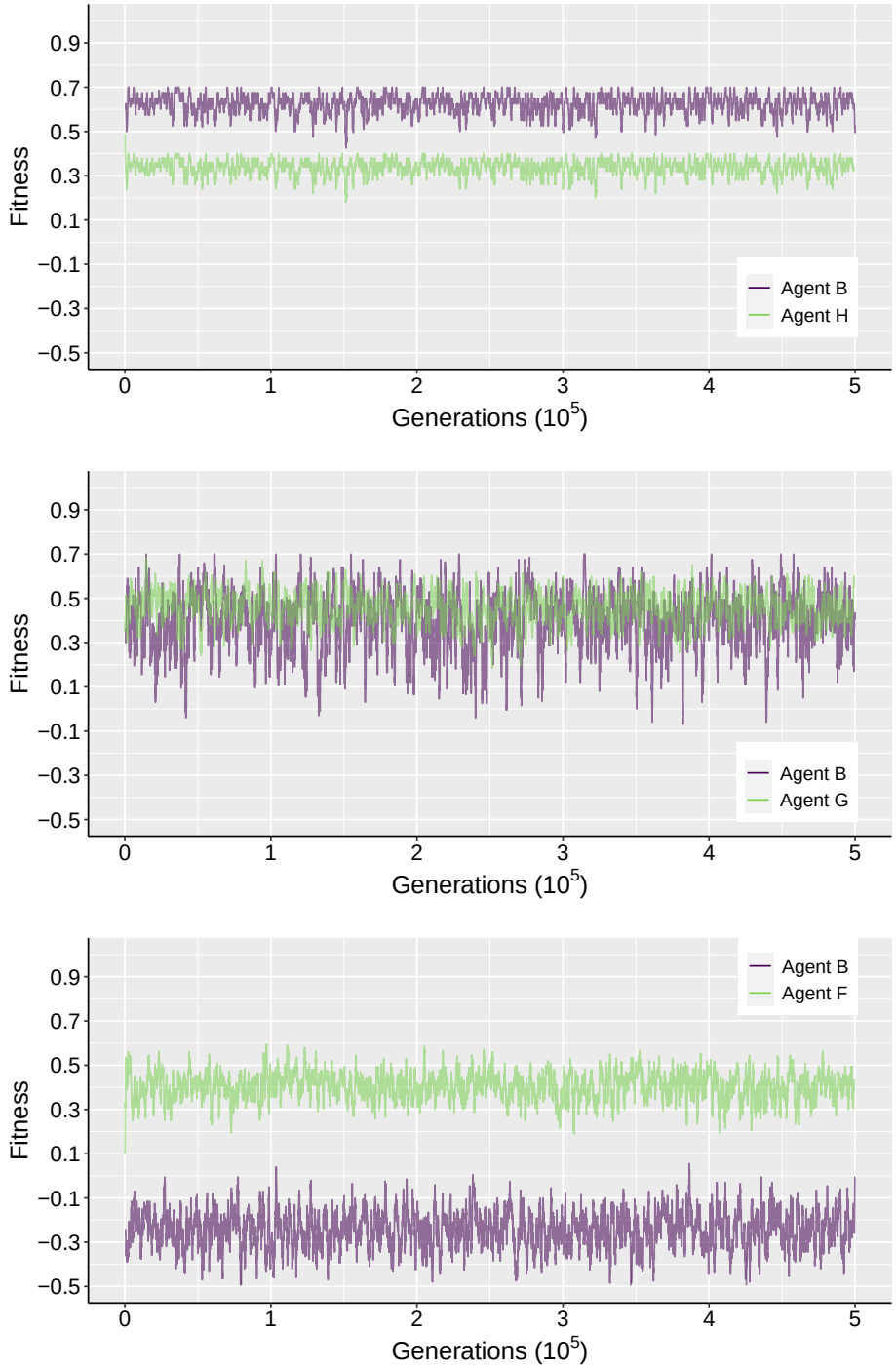


Figure 4.3: The moving average fitness of Agent B and Agents (a) H, (b) G and (c) F, without ongoing evolution, after an initial period of evolving alone. When socially situated with different pairs, Agent B (a) exploits Agent H to receive a higher average fitness from not exerting as much effort, (b) is more unpredictable than when alone, and is often unable to achieve its goal, and (c) cannot achieve its goal. In each experiment, Agent B's partner performs similarly to when it is alone.

river in an attempt to deal with the interference – simply because the agents are unable to respond appropriately to their environment changing in unknown ways. This cooperative and exploitative behaviour is *emergent* because agents do not possess the capacity to perceive others, their actions, or their intentions, so these behaviours cannot be *intended*.

The observed fitnesses in this shared environment differ to when the agents alone, as evolved behaviour is affected by the actions of the other agent. Interference from the actions of others has the potential to cause learnt knowledge and evolved behaviour to become unreliable; this results in goal-rational agents behaving differently, despite having the required knowledge to achieve their goals encoded within their genotype. As the results are dependent on the interactions with a specific partner, graphing an average of all ten experiments would mask the specific interactions. A representative sample of three experiments showing average population fitnesses are therefore shown in Figures 4.3a to 4.3c.

The fitnesses fluctuate based on the actions of both individuals at every generation, and as such appear very volatile. Emergent exploitation has no long-term implication here as agents do not evolve, so their genotype does not change; agents would continue to achieve their goals as before if they were again alone in an environment. The critical observation here is that merely sharing an environment is enough to change one's behaviour and ability to achieve goals; this is because interference can cause the agents' world view to change in unanticipated ways, thus affecting their ability to achieve goals.

4.4.3 Coexistence with Continued Evolution

Here, Agents A through J are arranged into 30 random pairs, to explore how interference affects coevolution in a shared environment. These 30 experiments evolve agents in a shared environment for 1,500,000 generations. Figures 4.4 to 4.8 depict the best fitness in each population for the named agents. Interference is observed to have the potential to affect evolved and reliable behaviour, and thus an agent's ability to solve tasks independently – even when agents do not explicitly interact with one another, or when they are not aware of each other. The most commonly observed emergent dynamic when agents continue to evolve together is exploitation: one agent exploits the other to receive a higher payoff at the other's expense. No pair of agents are observed to both maintain their ability to achieve goals alone – at least one agent is observed to change its evolved behaviour when experiencing interference.

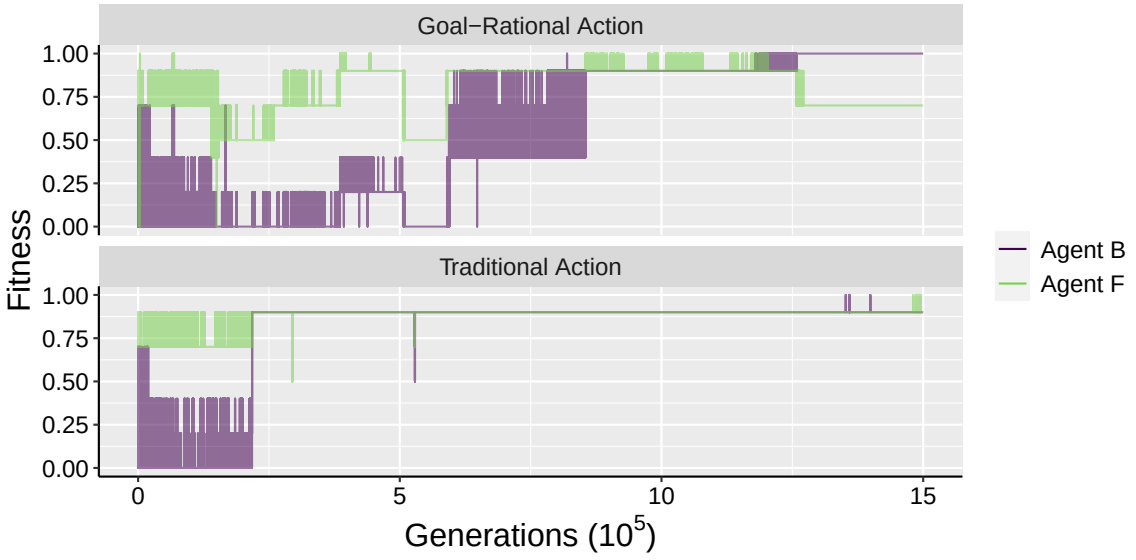


Figure 4.4: Agents B and F continuing to evolve together with goal-rational, and traditional action after a period of evolving alone. With goal-rational action, Agent B exploits Agent F and cannot achieve its goals alone; evolution is volatile as agents evolve to be codependent. Cooperation gives 0.9 fitness. If fitness < 0.7, agents do not achieve their goals. With traditional action, the agents endure a smaller period of low fitness, and get a better fitness of 0.9 by cooperating.

Figure 4.4 shows the most extreme case of interference observed; agents evolved to be codependent, as what agents learn depends on the actions of the other. There are four distinct 'phases' in evolution that can be observed in this graph. Firstly, between generations 0 and 500,000, Agent B loses its ability to achieve its goal, whilst Agent F is able to achieve a fitness of 0.7 or higher for the majority of the time. Secondly, between generations 500,000 and 900,000, Agent F mostly achieves a fitness of 0.9 or higher, whereas Agent B fluctuates between fitnesses of 0.4 (placing one Stone but only collecting one Resource), 0.7 (achieving the goal individually by placing two Stones) and 0.9 (cooperating with the other agent by only placing one Stone); sometimes Agent B cannot rely on the actions of Agent F in order to obtain a fitness of 0.9, so resorts to achieving its goal individually for the most-part. The third phase sees Agents B and F cooperating with each other to share the cost of building a bridge. In the fourth and final phase however, this cooperation diverges into exploitation, as Agent B begins to lose its ability to put Stones in the river and thus cooperate; reacting to this, Agent F evolves to achieve its goal individually, which Agent B exploits to achieve a higher fitness. While both agents endure periods of being unable to achieve their goal (depicted by a fitness below 0.7), Agent B is more negatively affected than Agent F; because of this volatility in evolution caused by interference, Agent B eventually evolves to exploit

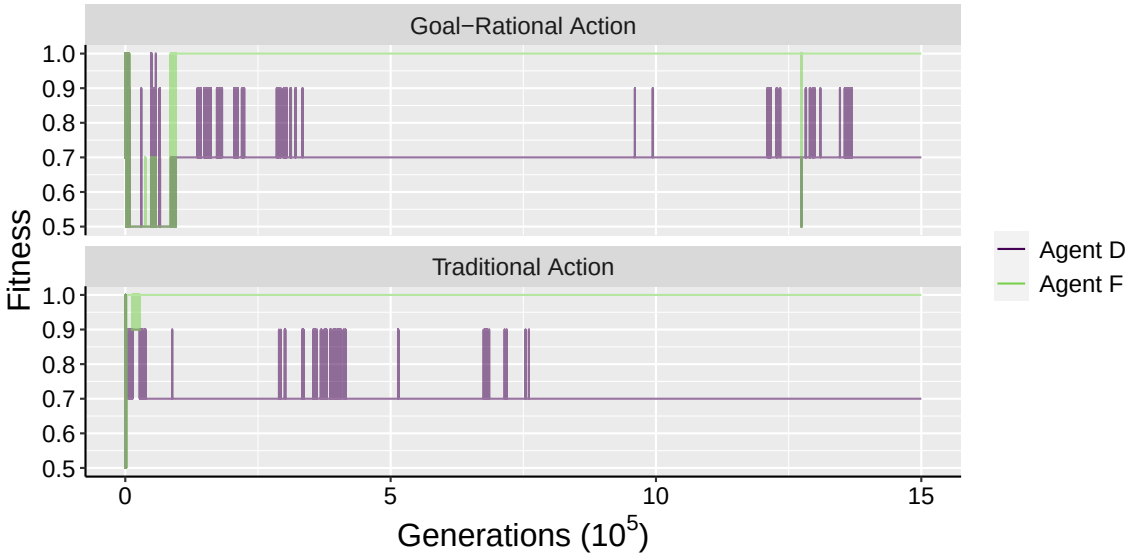


Figure 4.5: Agents D and F continuing to evolve together with goal-rational, and traditional action. With goal-rational action, both agents initially lose their ability to achieve goals, with a fitness of 0.5; Agent F then evolves to exploit Agent D. With traditional action, the agents endure less knowledge loss with traditional action and maintain their exploitative relationship in less time.

the other agent, and loses its ability to achieve its goal independently altogether. One can thus postulate that the effect of interference can be great, complex and uncertain as the world changes in unanticipated ways.

Other dynamics that can be observed to a lesser extent are periodic dips in fitness that can lead to exploitation (Figure 4.5). Early on in evolution, both Agents D and F are prone to exploiting the actions of the other; this exploitation is short-lived, as neither agent's actions are reliable enough for the other to consistently achieve their goals, which leads to a mutual loss of fitness and inability to achieve goals. Once Agent D becomes the first agent to re-evolve the goal-achieving behaviour of building a bridge individually, Agent F is then able to capitalise on this and an exploitative relationship emerges. The spikes observed in Agent D's fitness indicate that occasionally Agent F will cooperate, but is unreliable and predictably self-interested. As a result, Agent D sustains its independent behaviour, whilst Agent F evolves to capitalise on this to achieve a higher fitness.

Figure 4.6 shows a similar scenario between Agents D and G; there is a brief period of around 150,000 generations at the beginning of evolution where both agents are able to cooperate. However, Agent G tends to fluctuate between cooperative and exploitative behaviour, whereas Agent D consequently fluctuates between cooperation and achieving the goal individually. This is because the actions of Agent G are not predictable enough

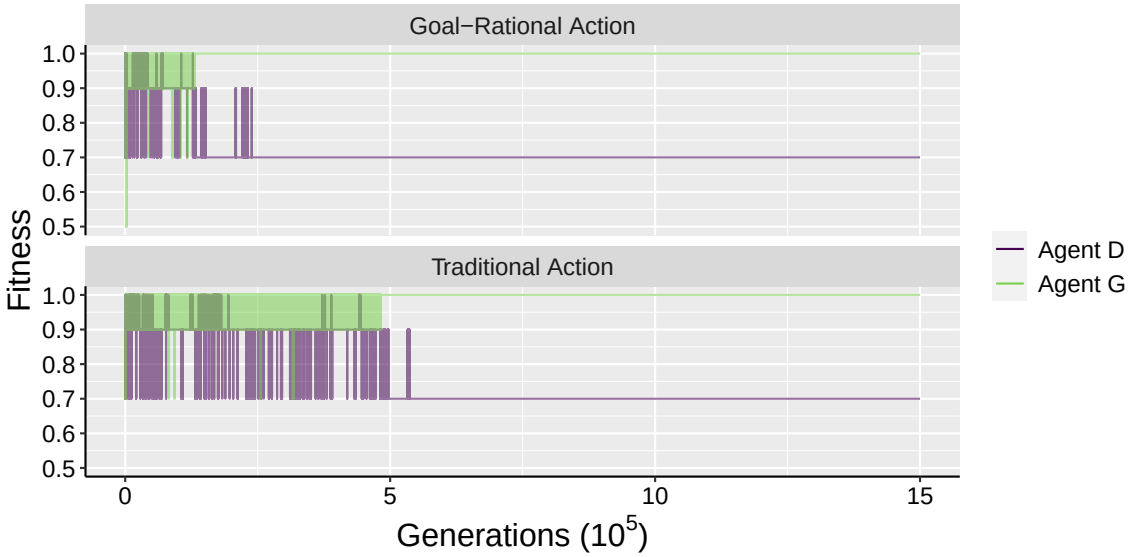


Figure 4.6: Agents D and G continuing to evolve together with goal-rational, and traditional action. With goal-rational action, Agent G cannot always rely on Agent D to exploit it, so it sometimes cooperates; Agent D cannot rely on G to cooperate, so evolves to achieve its goals alone.

to make cooperation beneficial for Agent D; eventually Agent D evolves to achieve its goals alone, whereas Agent G loses its ability to achieve its goals and exploits the other agent.

Figures 4.7 and 4.8 show exploitative relationships; the remainder of the 30 experiments are similar. As with Figure 4.5, peaks in the exploited agent's fitness indicate that the exploitative agent occasionally helps to build a bridge; it however evolves to be independent as the other agent is not predictable enough to rely on. This, again, evidences that interference affects how and what behaviours are evolved and maintained during evolution.

4.4.4 Implications of Coexistence

Evolved goal-achieving behaviour can potentially become unreliable when agents pursue individual goals with goal-rational action in shared environments. Interference from the actions of others changes each agent's perception of the world, and often leads to volatile evolution and knowledge loss; if agents cannot perceive the cause of the environmental changes, they will attempt to adapt and often lose learnt knowledge as a result. Exploitative agents for example depend on the other to achieve their goals, and thus 'forget' how to interact with Stones over time; this would be detrimental if they became suddenly alone in an environment, as they have evolved to be codependent rather than independent.

Further, the unanticipated and imperceptible actions of the other agent changes the

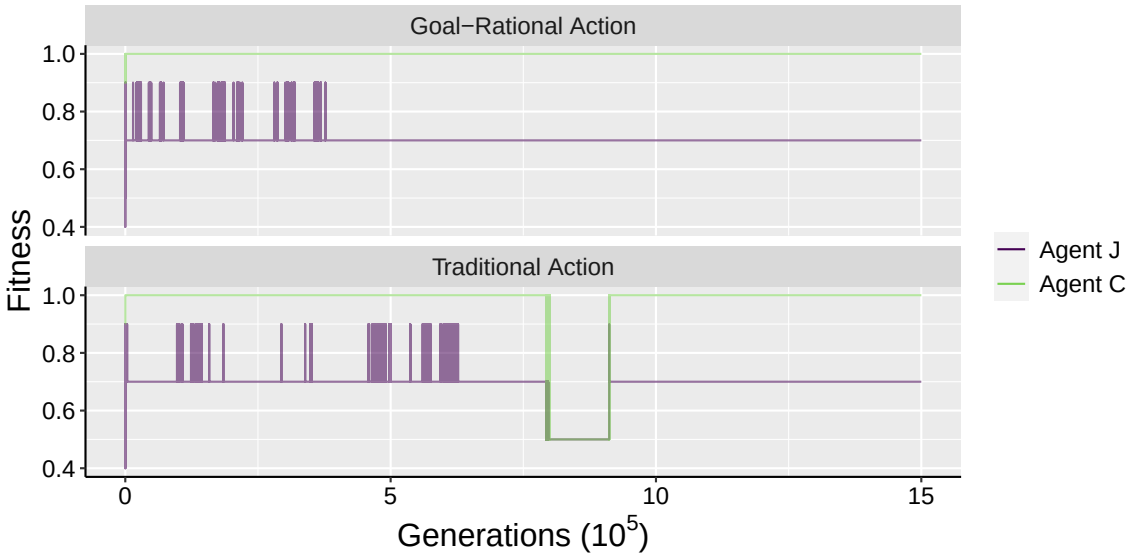


Figure 4.7: Agents J and C continuing to evolve together with goal-rational, and traditional action. With goal-rational action, Agent C exploits Agent J, which evolves to achieve goals alone. With traditional action, the agents endure low fitness temporarily with traditional action, but overcome this before maintaining an exploitative relationship.

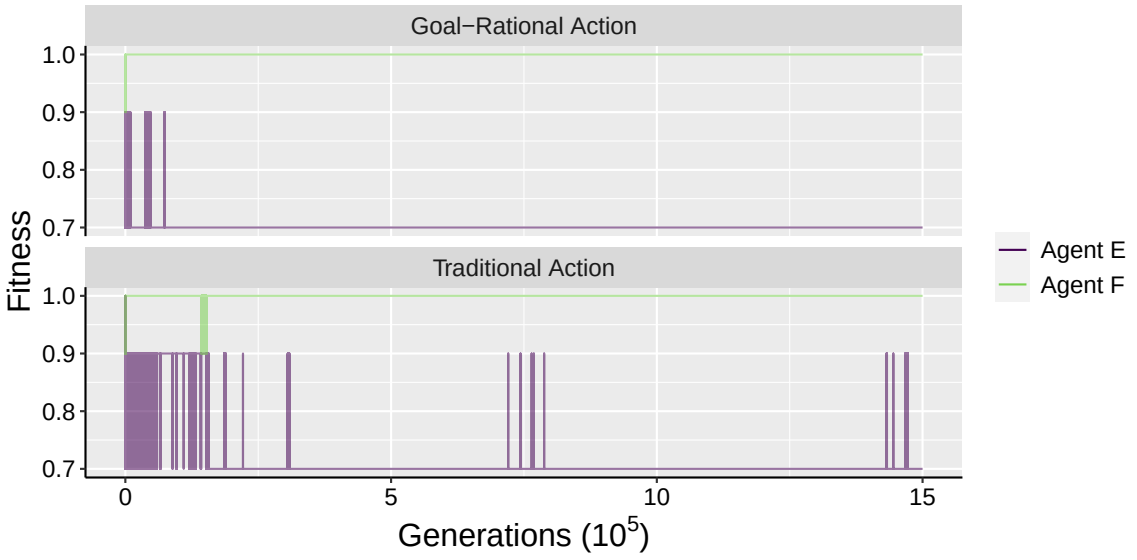


Figure 4.8: Agents E and F continuing to evolve together with goal-rational, and traditional action. With goal-rational action, Agent F exploits Agent E, which evolves to achieve goals alone. With traditional action, Agent F still evolves to exploit Agent E.

state of the environment, and can cause agents to alter their knowledge and behaviour in an attempt to adapt to the change in circumstances. This can lead to a change in the mapping of inputs to sub-goals in the deliberative network, and may result in drastic changes of behaviour. Specific examples could be an agent suddenly falling into the river due to an inability to respond to the environmental changes appropriately, or an agent that simply places many Stones into the river as its sub-goals have been altered to be repulsed from Resources. Interference can therefore impair an agent's ability to achieve its goals, even though it once possessed the behaviours required to be successful. An inability to overcome unexpected situations and unanticipated changes means agents are susceptible to knowledge loss and changes in behaviour. To combat this, traditional action is operationalised in an attempt to mitigate the effect of interference in socially situated agents.

4.4.5 Evolving with Traditional Action

Since goal-rational agents experience interference in shared environments, traditional action is hence compared to goal-rational action to explore how coexistence and interference may be affected by a different type of social action.

Figure 4.4 depicts the same pair of goal-rational agents with pure goal-rational action, and a blend of goal-rational and traditional action respectively. The volatility seen in the agents evolving with pure goal-rational action is drastically reduced with traditional action, as the fitness of each agent fluctuates less often. Further, the period of knowledge loss endured by Agent B as it learns about its new environment is much shorter with traditional action; additionally, both agents that use traditional action receive a mutual benefit from the presence of the other such that they evolve and maintain emergent cooperative behaviours. This enables both agents to achieve their goals with the best overall payoff.

Figure 4.5 presents a different pair of agents. With goal-rational action, both agents endure a period of low fitness at the beginning of evolution, as they are unable to respond appropriately to the interference they now face. In contrast, traditional action significantly reduces this initial period that agents are unable to achieve their goals; instead of a dip in fitness, agents briefly cooperate instead. Other than this, traditional action sees both agents evolve into the same exploitative relationship seen with just goal-rational action – in fewer generations. Traditional action is therefore observed to be beneficial overall for these agents.

Figure 4.6 demonstrates that cooperation can exist with exploitation when using traditional action. Fluctuations between fitnesses of 0.9 and 1.0 in Agent G indicate that it cannot rely completely on the actions of the other agent in order to exploit it, and therefore must still maintain knowledge of how to build a bridge. Agent D however is more independent, and fluctuates between a fitness of 0.7 and 0.9; Agent D cannot always rely on Agent G to cooperate, so maintains its ability to achieve its goal alone. Cooperation is maintained for longer with traditional action, so agents are therefore generally better off.

Figure 4.7 shows exploitative behaviour evolving in agents. However, employing traditional action causes a period of around 100,000 generations where both agents lose the ability to build bridges; this means that agents can only collect one Resource each, which gives a payoff of 0.5. As Agent C evolves to exploit Agent J, when Agent J's behaviour changes such that it cannot achieve its goal, neither can Agent C. This anomalous evolutionary event causes a temporary dip in fitness that is not observed without traditional action. However, traditional action enables the agents to overcome this dip in fitness, indicating that agents can potentially learn how to cope with unexpected events.

Further to the results discussed above, Figure 4.8 shows that traditional action can also have a minimal effect on evolution. In the remaining experiments, if pure goal-rational agents evolve to be exploitative, traditional agents will as well. Exploitative agents essentially 'forget' previously dependable knowledge that was encoded in their deliberative neural network, relying instead on the actions of the other agent to achieve their goal; this change in behaviour is simply caused by the interference that arises from coexistence.

4.4.6 Implications of Evolving with Traditional Action

Traditional action is seen in these results to have the potential to reduce the effect that interference can have on agent evolution and goal-achievement. Whilst intentional cooperation requires an explicit awareness of others [51], *unintentional*, emergent cooperation is observed in some experiments when agents receive a mutual benefit from the actions of the other; this enables the agents to pursue their own goals and achieve a better fitness overall. Additionally, coexisting agents are observed to evolve behaviours that can be maintained faster with traditional action, allowing them to recover from the unforeseen events or states caused by interference in shared environments.

4.5 Analysing the Effect of Social Action on Agent Evolution

Thus far, the effect of interference has been explored at the individual level, by examining how specific agents evolve in shared environments. As outlined in Section 4.3.3, the experiments in this section extend this investigation by observing how agents evolve more generally. Here, 100 agent populations are initialised with random weights and placed in either an environment in which they exist alone or with another agent, using three different types of social action: goal-rational, traditional, and random action.

4.5.1 Analysing the Effect of Traditional Action on Volatility

In these experiments, agents are evolved with goal-rational or traditional action, in order to observe the effect that the type of social action can have on evolutionary volatility in individual and social environments.

Analysing Fitness

The mean best-in-population fitness across all 100 agents in each experiment can be seen in Figure 4.9. The evolution of agents evolving with goal-rational or traditional action is the focus here, whereas evolving with random action is discussed in Section 4.5.2. One of the most noticeable differences between agents that evolve alone and those in shared environments is the increase in fitness that agents receive throughout evolution, when they are able to capitalise on the actions of others. When agents evolve alone, goal-rational action appears to enable agents to receive a marginally higher fitness than those that use traditional action; when agents share an environment, the opposite is true. In socially situated agents, traditional action can help agents receive a higher fitness faster; as evolution progresses, agents that use either goal-rational and traditional action begin to receive similar fitnesses. This indicates that traditional action is beneficial by helping agents to achieve goals early on in evolution, but the long-term effect is similar to goal-rational action.

Looking further into how the type of social action affects agent evolution, Table 4.2 presents a breakdown of the common goal-achieving behaviours that agents receive in each experiment after evolution. Over double the number of agents are able to achieve their goal when socially situated, compared to when they are alone; this is observed in agents that use goal-rational, and traditional action. Whilst fewer agents using traditional action are

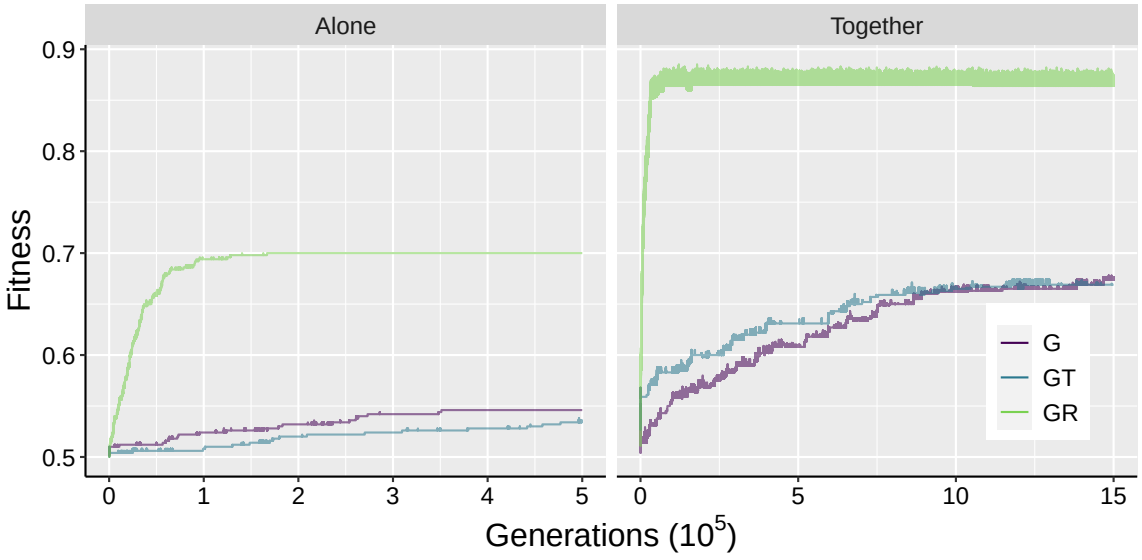


Figure 4.9: The mean best-in-population fitnesses of agents that evolve in an environment alone, and those that evolve in a shared environment with goal-rational (G), traditional (GT), or random (GR) action. *Note: the x-axes are not comparable because agents are evolved for 500,000 generations when alone, and 1,500,000 generations when together.*

able to achieve their goal than goal-rational agents both when socially situated and alone, some interesting observations can be made. Firstly, the same number of goal-rational agents achieve their goal individually in both single- and multi-agent environments; however, more agents that use traditional action are able to achieve their goal individually when sharing an environment. This is one indication that traditional action can help agents to overcome the negative effects of interference. Other than this, both goal-rational and traditional agents are able to exploit the actions of the other agent in the environment when socially situated, to receive a higher fitness. Unintentional, emergent cooperative behaviour is sometimes observed, but this is rare. These results would indicate that cooperation is a risky strategy when the agents do not possess the ability to perceive or reason about the existence of other agents, or the consequences of their actions. This would explain why agents seemingly prefer to either exploit the other agent or achieve their goal individually: because the actions of the other agent are unpredictable and therefore may not always be reliable, agents tend to resort to one strategy or the other, with cooperation observed infrequently.

Analysing Volatility

Over each of the 100 runs of this experiment, agents using traditional action are observed to have the same or a lower median SDoT, CACoT and CCoT than their purely goal-rational

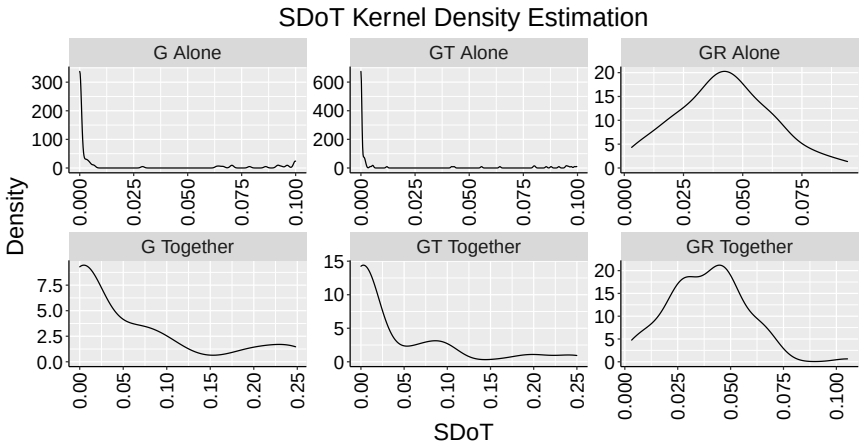
Experiment	Action	Fitness (% of Agents)				
		0.7	0.9	1.0	< 0.7	≥ 0.7
Alone	Goal-Rational	23	0	0	77	23
	Traditional	17	0	0	83	17
	Random	100	0	0	0	100
Together	Goal-Rational	23	5	22	50	50
	Traditional	20	1	25	54	46
	Random	42	4	54	0	100

Table 4.2: The percentage of agents that receive common fitnesses in each experiment, after 500,000 generations of evolving in an environment alone, or 1,500,000 generations of evolving in a shared environment, with goal-rational, traditional, or random action. 0.7 is a goal-achieving fitness after a bridge is built with two Stones; 0.9 is sharing the cost of bridge-building; 1.0 is exploitation; < 0.7 does not achieve the goal; ≥ 0.7 is a goal-achieving fitness.

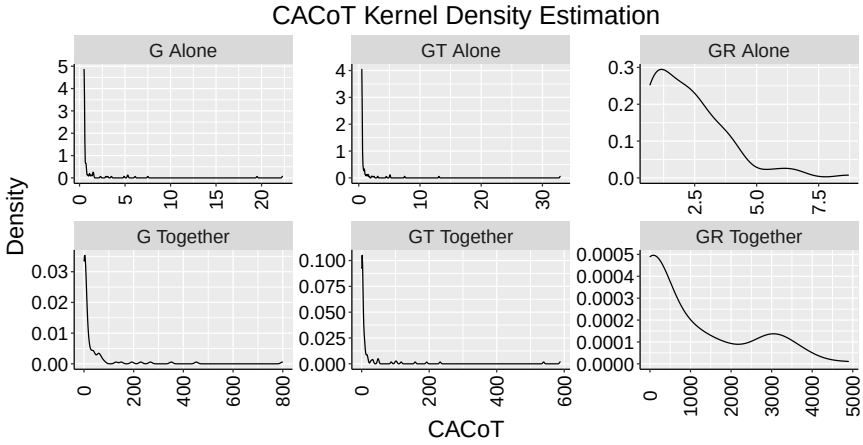
counterparts, which can be seen in Table 4.3. In other words, agents using traditional action typically have a lower volatility in evolution than purely goal-rational agents, where the fitness fluctuates less often and by smaller amounts during the course of evolution. The difference in volatility measured by the three metrics is much larger in socially situated agents than agents that are alone in an environment.

Agents using traditional action typically have an SDoT with less variability than those that do not (Table 4.3); this is more evident when agents exist alone, than when they are socially situated. Combined with a lower expected SDoT, and higher skew and excess positive kurtosis ($k_{excess} = k - 3$), agents using traditional action are expected to have a lower and more predictable SDoT. Kurtosis is used to analyse the tailedness and probability of outliers. These findings therefore indicate that traditional action enables agents to experience a lower, and more predictable standard deviation in fitness over time, with occasional outliers or extreme values; in other words, the fitness fluctuates less often in traditional agents during evolution, than in goal-rational agents. The shapes of these distributions can be seen in the kernel density estimation plots presented in Figure 4.10; each of the volatility distributions for agents evolving with either goal-rational or traditional action are skewed to the right, indicating positive skewness.

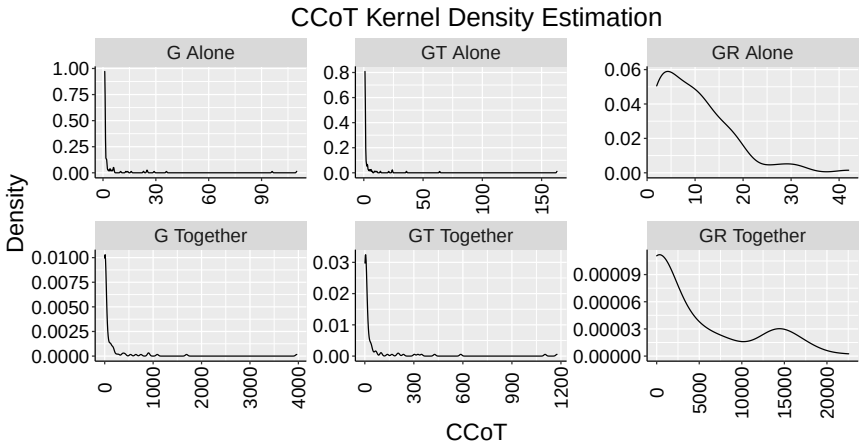
When alone, agents that use traditional action have a CACoT and CCoT with a higher variability and kurtosis than those that do not (Table 4.3). This changes when agents are socially situated, with traditional action reducing the variability and kurtosis in CACoT and CCoT. A lower mean CACoT and CCoT is observed in agents that use traditional action, than those that use goal-rational action, in both individual and shared environments. This



(a) SDoT Kernel Density



(b) CACoT Kernel Density



(c) CCoT Kernel Density

Figure 4.10: Kernel density estimation of the (a) SDoT, (b) CACoT and (c) CCoT of agents using goal-rational (G), traditional (GT), or random action (GR), evolved alone or together. *Note: the axes are not comparable due to the disparity in the densities of each distribution.*

Metric	Exp	Action	Moment				Median
			Mean	Variance	Skewness	Kurtosis	
SDoT	Alone	G	0.0155	0.001 08	1.84	4.56	0
		GT	0.0125	0.000 887	2.17	5.95	0
		GR	0.0422	0.000 394	0.171	2.74	0.0419
	Together	G	0.0589	0.006 21	1.29	3.35	0.0102
		GT	0.0424	0.004 74	1.75	5.04	0.003 32
		GR	0.0387	0.000 330	0.330	3.64	0.0406
CACoT	Alone	G	1.37	9.47	5.43	34.4	0.500
		GT	1.33	12.8	7.39	62.9	0.500
		GR	2.28	2.37	1.43	5.65	1.90
	Together	G	38.3	10 800	5.07	32.9	4.60
		GT	27.2	7420	5.19	31.6	2.70
		GR	1000	1 690 000	1.07	2.82	85.1
CCoT	Alone	G	5.37	237	5.43	34.4	1.00
		GT	5.15	321	7.39	62.9	1.00
		GR	9.88	59.1	1.43	5.65	8.00
	Together	G	150	215 000	6.22	48.3	10.0
		GT	64.4	33 000	4.63	26.2	7.50
		GR	4560	37 500 000	1.11	2.85	188

Table 4.3: Statistical moments and median (to 3 S.F.) of the SDoT, CACoT and CCoT of the best-in-population agents in each experiment (Exp) after 500,000 generations of evolving alone, or 1,500,000 generations together. Agents evolve with goal-rational (G), traditional (GT), or random (GR) action.

shows that traditional action can reduce volatility in evolution when agents experience interference. Agents that use traditional action typically have a lower CACoT than those that do not, indicated by a greater skewness; this is the same for the CCoT in agents that use traditional action when alone. The mean CCoT and skewness of the CCoT are lower in socially situated agents that use traditional action than those that do not; this means that traditional action typically reduces the fluctuations in fitness that agents experience during evolution, compared to goal-rational agents.

Statistical Tests

According to Yap and Sim [241], the Shapiro-Wilk normality test is powerful for distributions that are suspected to be asymmetric, as well as those that are symmetric with both high and low kurtosis values. As such, a Shapiro-Wilk normality test was conducted for each of the SDoT, CACoT and CCoT distributions for each experiment; the results of these tests can be found in Table B.1 in Appendix B.1. Results were significant for each with a p-value below 0.05, indicating that the distributions are non-normal. Wilcoxon Signed Rank tests were therefore used to compare the effect each action type had on evolution. These

are non-parametric tests used to compare the medians of two paired distributions; the null hypothesis for two-tailed tests is that the medians are identical, whereas one-tailed tests compare the directional difference in the distribution medians. One two-sided ($G \neq GT$) and two one-sided tests ($G > GT$, and $G < GT$) were conducted to compare the volatility experienced by agents that use either goal-rational or traditional action; the results of these statistical tests are presented in Table 4.4.

For each metric, no significant difference was found between the volatility of agents that evolve alone with goal-rational action, and those that evolve with traditional action. This is not surprising, as the agents do not experience interference and thus the environment is more predictable – resulting in similar levels of evolutionary volatility being observed despite the type of action employed. However, when agents are socially situated and experience interference, goal-rational agents have a significantly higher SDoT and CCoT ($G > GT$, $p < 0.05$, Table 4.4); traditional action in socially situated agents therefore reduces volatility in evolution compared to goal-rational action alone.

There is a very weak relationship between agents that evolve alone with goal-rational action, and those that evolve with traditional action. Whilst no significant difference between the volatility of agents employing these two types of action was found, the effect size estimate r for each pairing is positive, which indicates that goal-rational agents experience more volatility than agents using traditional action (Table 4.4). These effect sizes are very small according to the boundaries outlined by Cohen [58] ($r = 0.1$ is said to be a 'small' effect), but is observable nonetheless. When agents are in a shared environment however, a 'small', positive effect size is observed; this indicates that traditional action has a stronger effect on socially situated agents, in terms of the amount of volatility experienced during evolution, than those in isolation.

4.5.2 Random Action: High Expected Fitness, High Volatility

As outlined in Section 4.3.3, random action is also introduced to ascertain whether traditional action is quantitatively different to adding Random Immigrants into the population. Here, the difference between random and traditional action when combined with goal-rational action is compared.

Metric	Experiment	Statistical Test Alternative Hypothesis			z	r
		$G \neq GT$	$G < GT$	$G > GT$		
SDoT	Alone	4.282×10^{-1}	7.885×10^{-1}	2.141×10^{-1}	0.7970	0.079 70
	Together	4.565×10^{-2} *	9.774×10^{-1}	2.283×10^{-2} *	2.000	0.2000 (S)
CACoT	Alone	7.669×10^{-1}	6.200×10^{-1}	3.835×10^{-1}	0.3010	0.030 10
	Together	1.920×10^{-1}	9.046×10^{-1}	9.601×10^{-2}	1.310	0.1310 (S)
CCoT	Alone	7.426×10^{-1}	6.322×10^{-1}	3.713×10^{-1}	0.3330	0.033 30
	Together	9.780×10^{-2}	9.515×10^{-1}	4.890×10^{-2} *	1.660	0.1660 (S)

Table 4.4: Wilcoxon Signed Rank statistical tests comparing the volatility metrics (Section 5.5.3) of agents evolving alone or together, with goal-rational (G) or traditional (GT) action. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

Analysing Fitness

Referring back to Figure 4.9, the effect of random action on evolution compared to either goal-rational or traditional action is immediately obvious: agents using random action are able to receive higher fitnesses during evolution than those that use the other two types of action. This effect is so pronounced, that 100% of agents are able to achieve their goal with random action in both single- and multi-agent environments in the RCD (Table 4.2). Similarly to the discussion of individual, cooperative and exploitative goal-achieving behaviour in Section 4.5.1, agents that use random action in shared environments tend to either achieve their goal individually, or exploit the other agent. Whilst this shows that random action is beneficial for agents in terms of the fitness received during evolution, Figure 4.9 shows noticeably more fluctuations in fitness over time in agents that evolve in shared environments. This volatility will consequently be analysed in closer detail.

Analysing Volatility

Goal-rational agents that use random action have a higher mean SDoT, CACoT and CCoT than those that use the other two types of action. The exception to this is that the expected SDoT is lower when agents experience interference with random action, despite the median still being larger than the other types of action (Table 4.3); a lower mean SDoT but higher mean CACoT and CCoT indicates that agents using random action in shared environments will predictably experience high levels of volatility and fluctuations in fitness during evolution. Additionally, agents that use random action have an SDoT with lower variability, and those that act alone have a CACoT and CCoT with lower variability than those that use

goal-rational or traditional action. Therefore, when agents do not experience interference, using random action means the SDoT, CACoT and CCoT will be predictably high due to low dispersion around the mean. This changes drastically when agents are socially situated (Table 4.3), as interference can cause fitness in agents using random action to fluctuate often, indicated by the variance in the CACoT and CCoT being orders of magnitude higher than in agents that use either goal-rational, or traditional and goal-rational action.

In all experiments, the SDoT, CACoT and CCoT are positively skewed, with random action being the least skewed; this means that the CACoT and CCoT are typically higher in agents that use random action compared to the other types of action, as the values are more symmetrical around the higher means (Table 4.3). It can therefore be said that evolution with random action is more volatile than goal-rational or traditional action.

The SDoT, CACoT and CCoT in both goal-rational and traditional action are highly leptokurtic, meaning there is positive excess kurtosis; in other words, these distributions have 'fat' tails, which indicates that extreme outliers will be common (Table 4.3). The SDoT, CACoT and CCoT are expected to be less extreme in agents using random action as the kurtosis is lower than in agents using the other two types of action; saying this, whilst outliers will be less common and less extreme in agents that use random action, the SDoT, CACoT and CCoT are generally higher overall.

Statistical Tests

A Shapiro-Wilk normality test was also conducted for the volatility distributions of agents using random action, when they in environments either alone or with another; the results of these tests are presented in Table B.1 in Appendix B.1. The results of these tests were significant at $p < 0.05$, indicating that the CACoT and CCoT distributions for these agents are non-normal; however, the SDoT of agents evolving with random action cannot be said to be non-normal, as the p-value is greater than 0.05. This is caused by the high number of agents that achieve their goal, therefore reducing the variability in fitness.

Consequently, Wilcoxon Signed Rank statistical tests were used to ascertain whether there is any observable difference between the volatility experienced by agents that use random action, and those that either use goal-rational action (G compared with GR), or traditional action (GT compared with GR). Wilcoxon Signed Rank tests can still be used for the SDoT distributions as this type of test does not assume anything about the

Metric	Experiment	Statistical Test Alternative Hypothesis			z	r	
		$G \neq GR$	$G < GR$	$G > GR$			
SDoT	Alone	1.415×10^{-8} *	7.077×10^{-9} *	1.000	−5.670	−0.5670	(L)
	Together	4.993×10^{-1}	7.515×10^{-1}	2.496×10^{-1}	0.6770	0.067 70	
CACoT	Alone	6.814×10^{-9} *	3.407×10^{-9} *	1.000	−5.800	−0.5800	(L)
	Together	6.350×10^{-12}*	3.177×10^{-12}*	1.000	−6.870	−0.6870	(L)
CCoT	Alone	6.146×10^{-9} *	3.073×10^{-8} *	1.000	−5.810	−0.5810	(L)
	Together	5.557×10^{-11}*	2.778×10^{-11}*	1.000	−6.560	−0.6560	(L)

Table 4.5: Wilcoxon Signed Rank statistical tests comparing the volatility metrics (Section 5.5.3) of agents evolving alone or together, with goal-rational (G) or random (GR) action. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

distribution (such as normality). The results of these statistical tests are presented in Tables 4.5 and 4.6. Agents that use random action have a significantly higher mean CACoT and CCoT than those that use the other two types of action, as well as a significantly higher mean SDoT when agents are alone ($p < 0.05$). This indicates that the evolution of agents using random action is significantly different to the other types of action, and agents are affected significantly more by interference. The exception to these observations of statistical significance, is that the SDoT of socially situated agents evolving with random action is not found to be different to that of agents evolving with either goal-rational or traditional action. Upon looking at the data regarding volatility presented in Tables 4.3, 4.5 and 4.6, it would seem that the magnitude and frequency of the fluctuations in fitness over time (captured with the CACoT and CCoT metrics respectively) for socially situated agents using random action are significantly higher than in those that use the other two types of action; however, the SDoT remains low (without a statistically significant difference to agents using the other two types of action), which indicates that these agents are predictably, and highly volatile when sharing an environment with another agent.

To demonstrate this further, a 'large', negative effect size [58] is observed in each pairing, which indicates that random action has a strong effect on the amount of volatility that agents experience compared to goal-rational agents. Aligning with the exception discussed above, random action has a negligible effect on the SDoT of agents that evolve together compared to goal-rational action (Table 4.5); in fact, as the effect size estimate r is positive, this indicates that goal-rational agents experience more volatility than agents using random action here, which might be caused by the latter agents receiving predictably high fitnesses

Metric	Experiment	Statistical Test Alternative Hypothesis			z	r
		$GT \neq GR$	$GT < GR$	$GT > GR$		
SDoT	Alone	$2.643\times10^{-11*}$	$1.322\times10^{-11*}$	1.000	−6.670	−0.6670 (L)
	Together	1.816×10^{-1}	9.081×10^{-2}	9.098×10^{-1}	−1.340	−0.1340 (S)
CACoT	Alone	$9.769\times10^{-11*}$	$4.884\times10^{-11*}$	1.000	−6.470	−0.6470 (L)
	Together	$1.797\times10^{-15*}$	$8.986\times10^{-16*}$	1.000	−7.960	−0.7960 (L)
CCoT	Alone	$9.594\times10^{-11*}$	$4.797\times10^{-11*}$	1.000	−6.480	−0.6480 (L)
	Together	$1.473\times10^{-14*}$	$7.363\times10^{-15*}$	1.000	−7.690	−0.7690 (L)

Table 4.6: Wilcoxon Signed Rank statistical tests comparing the volatility metrics (Section 5.5.3) of agents evolving alone or together, with traditional (GT) or random (GR) action. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

with predictably high volatility. Additionally, a 'small', negative effect size [58] is observed between the SDoT of traditional agents and those that use random action, which indicates that again, random action increases the volatility that agents experience.

To summarise, agents that use random action typically have a significantly higher SDoT, CACoT and CCoT during evolution than those that use traditional or pure goal-rational action; this is due to a higher mean and median, combined with less skewness and kurtosis. When agents exist alone, the SDoT, CACoT and CCoT have less variability, but in social environments, the variance is extremely large. This indicates that evolution is more volatile when agents experience interference, than when they are situated in an environment alone. The shapes of these distributions can be seen in Figure 4.10, which are kernel density estimation plots for each distribution. For agents using random action, each of the volatility distributions are skewed to the right as there is positive skew; there is noticeably more variability in the volatility experienced by agents using random action than those using the other two types of action however, as the data is more dispersed. It can therefore be concluded that evolution with random action is extremely volatile compared to the other types of actions implemented, especially in shared environments.

4.5.3 Balancing the Fitness and Volatility Trade-Off

In this section, the fitness of agents, as well as the evolutionary volatility they experience is considered for each of the three types of action operationalised in this study: goal-rational, traditional, and random action. Firstly, Table 4.7 presents the statistical moments for the best-in-population fitness that agents across all experiments in this study receive. Secondly,

Experiment	Action	Moment				Median
		Mean	Variance	Skewness	Kurtosis	
Alone	Goal-Rational	0.546	0.007 16	1.28	2.65	0.500
	Traditional	0.534	0.005 70	1.76	4.09	0.500
	Random	0.700	0	NaN	NaN	0.700
Together	Goal-Rational	0.676	0.0416	0.636	1.78	0.600
	Traditional	0.669	0.0440	0.725	1.83	0.500
	Random	0.870	0.0215	−0.282	1.11	1.00

Table 4.7: Statistical moments and median (to 3 S.F.) of the best-in-population fitness after 500,000 generations of evolving alone, or 1,500,000 generations evolving together, with goal-rational, traditional, or random action.

a Shapiro-Wilk test for normality was conducted for the fitness distributions for agents using each type of action, in both single- and multi-agent experiments; the results for these tests can be found in Table B.2 in Appendix B.1. These results were significant at $p < 0.05$, which indicates that the fitness distributions for each experiment are non-normal. The exception to this is observed in agents that evolve alone with random action because 100% of the agents receive a fitness of 0.7; as such, the Shapiro-Wilk test cannot be conducted for this set of results as all the values are identical. After conducting the Shapiro-Wilk tests for normality, Wilcoxon Signed Rank statistical tests were then conducted to compare each type of action; the results of these tests are presented in Table 4.8.

In each experiment, agents that use random action receive a significantly higher fitness than agents using either goal-rational or traditional action ($G < GR$ and $GT < GR$, $p < 0.05$, Table 4.8). The shapes of these distributions can be seen in the kernel density estimation plots presented in Figure 4.11; the fitness distributions for agents evolving with either goal-rational or traditional action are positively skewed, whereas the distribution for socially situated agents evolving with random action is negatively skewed. The effect of random action is so great that all 100 agents that use random action are able to achieve their goal while existing in an environment alone (Table 4.2); this is visible in Figure 4.11, as the peak of the distribution is at a fitness of 0.7 – which is achieved by all agents in the experiment. As a result, there is no skew or kurtosis for this distribution (Table 4.7), as all values in the distribution are identical. In shared environments however, random action allows all 100 agents to achieve their goal (Table 4.2), but a median fitness of 1.0 indicates that these agents are expected to evolve exploitative behaviour. Whilst this is beneficial fitness-wise, the agents would most likely lose the ability to achieve their goals

Action (x-y)	Experiment	Statistical Test Alternative Hypothesis			z	r	
		$x \neq y$	$x < y$	$x > y$			
G-GT	Alone	3.217×10^{-1}	8.435×10^{-1}	1.608×10^{-1}	1.000	0.1000	(S)
	Together	8.379×10^{-1}	5.836×10^{-1}	4.190×10^{-1}	0.2080	0.020 80	
G-GR	Alone	2.200×10^{-16}*	2.200×10^{-16}*	1.000	−8.770	−0.8770	(L)
	Together	2.453×10^{-9} *	1.226×10^{-9} *	1.000	−5.970	−0.5970	(L)
GT-GR	Alone	2.200×10^{-16}*	2.200×10^{-16}*	1.000	−9.110	−0.9110	(L)
	Together	4.243×10^{-9} *	2.122×10^{-9} *	1.000	−5.880	−0.5880	(L)

Table 4.8: Wilcoxon Signed Rank statistical tests comparing the fitness of agents that use goal-rational (G), traditional (GT), or random (GR) action, where agents evolve alone or together with another. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

independently (or not evolve this ability to begin with), and not maintain bridge-building behaviours either. This would be problematic if these agents were to find themselves in an environment alone, as they would have to evolve this goal-achieving behaviour from scratch. Although this higher expected fitness is desirable, dependency on others would not be desirable if the existence of others is unpredictable, or the environment is extremely dynamic; this is because there may be long periods where the agent is unable to achieve its goals.

Table 4.8 also shows that a 'large', negative effect size estimate r [58] is recorded when comparing random action to both goal-rational and traditional action; this shows that random action has a large effect on the fitness that agents receive, and that agents using random action receive higher fitnesses than those that use the other two types of action. In comparison, only a 'small' effect size is observed when comparing goal-rational and traditional action in agents that evolve alone, and a marginal effect is observed in those that evolve together. This means that goal-rational action has a small benefit over traditional action when agents evolve alone (in terms of the fitness received), but this is marginal when agents exist in a shared environment. These results show that there is little disadvantage for using traditional action over goal-rational action, because agents will receive similar fitnesses regardless of the type of action they use.

While the data presented in Table 4.7 shows that agents using pure goal-rational action receive a mean fitness marginally higher than those that use traditional action, no statistically significant difference is found between the two (Table 4.8); this indicates that the fitness agents receive is similar when using either goal-rational or traditional action.

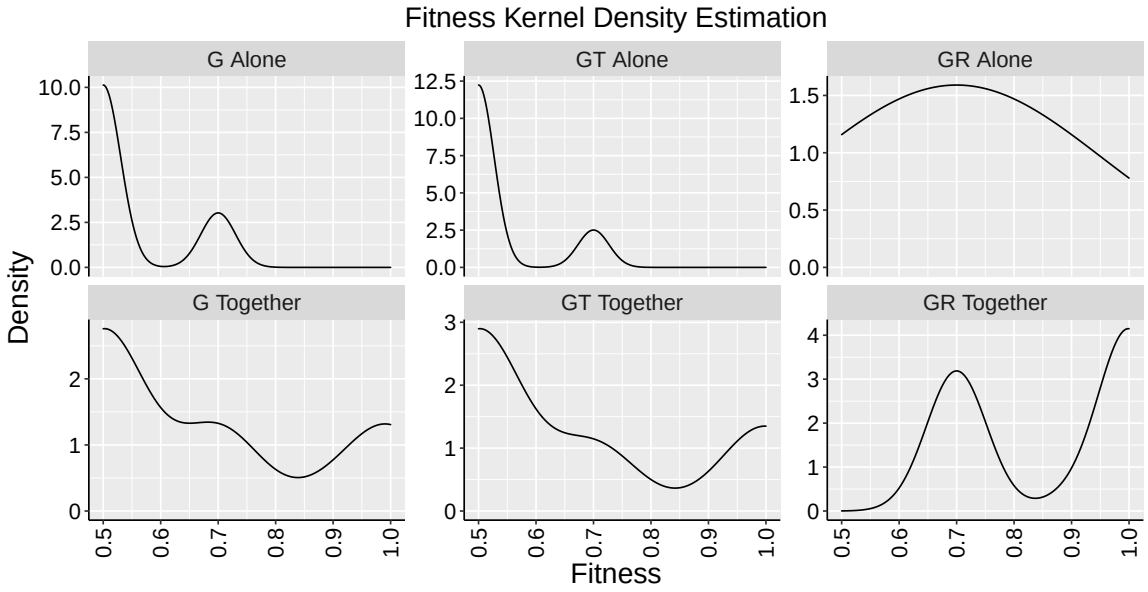


Figure 4.11: Kernel density estimation of the fitness of agents using goal-rational (G), traditional (GT), or random (GR) action, evolved alone or together. *Note: the y-axes are not comparable due to the disparity in the densities of each distribution.*

4.5.4 Analysing Goal-Achievement over Evolution

In addition to analysing the fitness that agents receive with each type of action at the end of evolution, analysing how often these agents achieve their goals during evolution can also provide useful insight into how beneficial each type of action is for agents more generally.

Figure 4.12 presents a box plot of the number of generations that agents receive a goal-achieving fitness (≥ 0.7) with each type of action, when they evolve alone for 500,000 generations or for 1,500,000 generations when socially situated. It is clear from this figure that random action enables agents to achieve their goals for more generations throughout evolution than either goal-rational or traditional action – in both individual and shared environments. This corroborates the analysis presented above, which concludes that random action enables agents to achieve their goals more often than the other two types of action, but at the cost of evolutionary volatility.

Goal-rational and traditional agents acting in an environment alone appear to be similarly successful (in terms of the number of generations in which goals are achieved) when looking at Figure 4.12. Despite this, these agents will predictably only achieve their goals for a small portion of evolution, as evidenced by the first and third quartiles of the box plots being in narrow range of each other. The range of successful generations however is

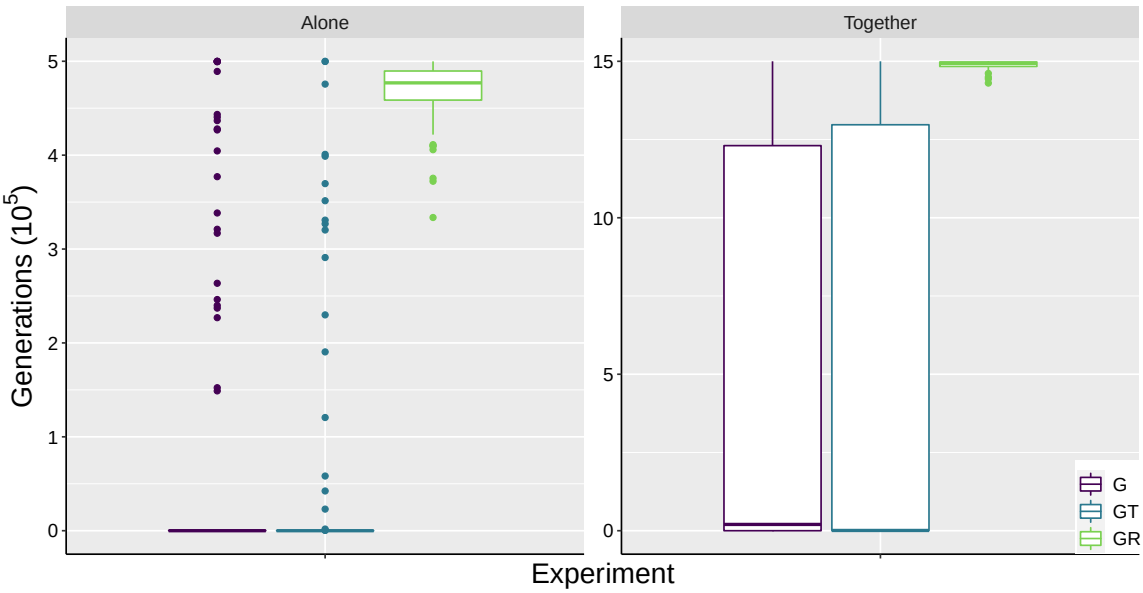


Figure 4.12: Box plot depicting the number of generations agents receive a goal-achieving fitness (≥ 0.7) during 500,000 generations of evolving alone, or 1,500,000 generations of evolving together, with either goal-rational (G), traditional (GT), or random (GR) action.

much greater in agents that evolve in shared environments compared to those that evolve in isolation; this demonstrates that existing in a shared environment allows agents to be more successful and achieve their goals more often compared to when agents exist alone – despite the fact that the agents in these experiments are unaware of the presence of others or the impact of their actions.

In order to study this in more detail, Table 4.9 presents the statistics for the box plots depicted in Figure 4.12, along with the mean number of generations that agents in each experiment achieve their goals. The data shows that goal-rational action enables agents to achieve their goals more often during evolution than traditional action, and that random action improves this further; this difference is more prominent when agents evolve alone than together. Figure 4.9 shows that the mean best-in-population fitness is higher for goal-rational agents throughout evolution compared to traditional agents when they act alone, however the reverse is true when agents are socially situated; further, traditional action enables agents to discover goal-achieving behaviour earlier in the evolutionary process than goal-rational action in shared environments. This explains why there is less of a difference between the number of generations that agents are successful when evolving together, and is further evidence that traditional action can be beneficial for socially situated agents.

To ascertain the extent to which the type of action influences the number of generations

Experiment	Action	Min	Q_1	Median	Mean	Q_3	Max
Alone	Goal-Rational	1	1	1	84 378	258	499 959
	Traditional	1	1	1	49 326	30	499 939
	Random	333 599	458 627	476 986	469 216	489 607	499 880
Together	Goal-Rational	1	17	20 035	562 381	1 230 485	1 499 981
	Traditional	1	15	801	550 338	1 297 230	1 499 949
	Random	1 430 421	1 483 553	1 493 578	1 488 840	1 497 800	1 499 987

Table 4.9: The minimum and maximum number of generations that agents achieve their goal in each experiment, along with the first, second and third quartiles (Q_1, median, and Q_3 respectively), and the mean. Agents evolve with either goal-rational, traditional, or random action, in an environment alone or together with another. Values are rounded to the nearest whole number.

that agents are able to achieve their goals throughout evolution, statistical tests need to be conducted as well. Shapiro-Wilk tests for normality indicated that each distribution was non-normal; the results of these tests can be found in Table B.3 in Appendix B.1. Consequently, Wilcoxon Signed Rank statistical tests were chosen as a suitable test for this purpose, and the results of which are presented in Table 4.10. Random action was found to be significantly different to both goal-rational and traditional action ($p < 0.05$). Further, a significant directional difference between the medians of random action, and goal-rational and traditional action, indicates that random action enables agents to achieve their goals more often during evolution than the other two types of actions ($p < 0.5$, $x < y$). Table 4.10 also shows that random action has a 'large' effect [58] on the number of generations that agents successfully achieve their goals, compared to the other two types of action. Comparatively, no significant difference was found between goal-rational and traditional agents, in terms of the number of generations in which they achieve their goals during evolution. This means that despite goal-rational action appearing at first to be more beneficial to agents than traditional action when analysing the first, second and third quartiles of the distributions (Table 4.9), there is no discernible difference between the two types of action in terms of goal-achievement during evolution. Goal-rational action has a 'small' effect [58] on the number of generations that agents evolving in isolation achieve their goals (Table 4.10); however, only a marginal effect is observed when agents exist in an environment with another. This shows that there is little difference between the number of generations that traditional agents and goal-rational agents are successful during evolution – with socially situated agents being more similar than those in isolation. Combined with the analysis in earlier sections, traditional action can therefore be said to enable agents to perform similarly to goal-rational action in terms of the number of generations agents

Action (x-y)	Experiment	Statistical Test Alternative Hypothesis			z	r	
		$x \neq y$	$x < y$	$x > y$			
G-GT	Alone	1.453×10^{-1}	9.286×10^{-1}	7.265×10^{-2}	1.460	0.1460	(S)
	Together	9.403×10^{-1}	5.313×10^{-1}	4.701×10^{-1}	0.076 70	0.007 670	
G-GR	Alone	$2.220\times10^{-16*}$	$2.220\times10^{-16*}$	1.000	−8.560	−0.8560	(L)
	Together	$2.220\times10^{-16*}$	$2.220\times10^{-16*}$	1.000	−8.620	−0.8620	(L)
GT-GR	Alone	$2.220\times10^{-16*}$	$2.220\times10^{-16*}$	1.000	−8.660	−0.8660	(L)
	Together	$2.220\times10^{-16*}$	$2.220\times10^{-16*}$	1.000	−8.290	−0.8290	(L)

Table 4.10: Wilcoxon Signed Rank statistical tests comparing the number of generations that the best-in-population agents receive a goal-achieving fitness (≥ 0.7) in each experiment, where agents use goal-rational (G), traditional (GT), or random (GR) action, and evolve alone or together with another. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

achieve their goals during evolution, as well as the number of agents that achieve their goals at the end of evolution. Furthermore, whilst there is no difference observed between agents that use goal-rational and traditional action in terms of the fitness received, the volatility these agents experience during evolution is significantly reduced with traditional action.

Drawing on the analysis conducted in this chapter, there appears to be a trade-off between the expected fitness and the evolutionary volatility of socially situated agents. Whilst random action may seem like the obvious choice for agents in shared environments because of this increase in fitness, an inability to achieve goals individually may be detrimental to the overall evolution of the agent if the actions of another agent cannot always be depended upon. A median fitness of 1.0 in socially situated agents that use random action implies that over half of the agents are able to exploit the other agent in the environment to achieve the goal without incurring a cost for bridge-building; this also means that in over half of these experiments, the partner agent has evolved the ability to build a bridge independently. In environments where the other agent is not always dependable, or is prone to change, agents using random action would experience high levels of volatility, and often periods of being unable to achieve their goal. Random action therefore significantly increases the fitness of agents, but at the cost of evolutionary volatility: this high fitness fluctuates often, and by large amounts, during the course of evolution.

In contrast, agents using goal-rational and traditional action experience significantly less volatility overall than agents using random action – at the cost of receiving a lower fitness overall. When comparing goal-rational to traditional action however, the addition

of traditional action significantly reduces the volatility in evolution in otherwise purely goal-rational agents, without sacrificing fitness. It can therefore be said that incorporating traditional action into otherwise goal-rational agents is a logical way to reduce evolutionary volatility in agents when they experience interference from the actions of other agents in the environment.

4.6 Discussion and Implications

As real-world sociotechnical systems grow in size, the components that these large, complex systems are composed of will be more likely to interact with one another – intentionally or not [101]. As a result, it is important to design these systems with the ability to achieve their goals, despite the social nature of their environments. Moreover, since it is not possible for a system component to know of all other components or actors within the environment (especially with the increasing size of said systems) [112], it might not always be possible to make decisions with an explicit awareness of other actors that have the potential to interfere with one's own ability to operate and pursue goals.

The results presented within this chapter demonstrate the effect that sharing an environment with other, unknown actors can have on the evolution of artificial agents, and their ability to achieve individual goals. Existing in a shared environment can be beneficial to agents, as cooperative or exploitative behaviour can lead to a higher fitness than is accessible when alone. However, whilst this may be useful when the presence and behaviour of the unknown other agent is reliable, if the environment suddenly becomes solitary, then the agent may have evolved to be dependent on the actions of the other agent in order to achieve its goals. In the real world, the implications of one system or component depending on the actions of another in order to achieve its goals could prove to be catastrophic; in the case that the exploited system is taken offline, or changes its behaviour, the exploitative system may no longer be able to function correctly. Further, a lack of knowledge of the existence of the exploited system, and consequently that its actions are enabling the exploitative system to succeed in the first place, would mean that the exploitative system would have to relearn how to achieve its goals on its own. It may not be possible for these systems to obtain information about others in their environment that can potentially affect them, which calls for exploration into how systems can be designed to operate in shared

environments to mitigate interference from others *without* such knowledge.

By acting in a more socially-oriented manner, agents in this study are shown to experience less evolutionary volatility while receiving a similar fitness to agents that act in a goal-oriented manner. This small change of adding traditional action to the evolutionary process enables agents to behave in a more predictable manner (i.e. their fitness changes less frequently over time); crucially, this benefit is observed in agents that have no knowledge of whether they are in a shared or isolated environment. As a result, acting in a more socially-oriented way in general – regardless of whether other agents exist in the environment that are known or otherwise – can help agents to mitigate the effects of interference without explicitly being aware of it. This would be desirable in the socially intelligent systems discussed by Bellman et al. [28] and Castelfranchi [51], which would be capable of integrating themselves into shared environments with heterogeneous actors, as well as operating predictably in order to achieve their goals. In contrast, agents in this study that use random action in this study employ a 'high risk, high reward' strategy, allowing them to achieve their goals more often than both goal-rational and traditional agents – but at the cost of evolutionary volatility. This could be useful in real-world scenarios where the best solution is valued despite the cost, however reliability and predictability in performance will likely be preferred. This is especially true in the case of systems that are critical to the functioning of society, such as traffic and waste water management systems [46]. Traditional action is therefore demonstrated to be one way in which systems can mitigate the effects of interference, without having to sacrifice performance (in this case, fitness) or reliability (in this case, volatility), and without requiring extensive knowledge about others in order to overcome the effects of interference from others within the environment.

4.7 Conclusion

The experimental study presented in this chapter explores the effect that interference can have on agent evolution and volatility, by evolving agents that either exist alone, or share an environment with another agent in the River Crossing Dilemma testbed. Humans have evolved to succeed in social environments by orienting their actions towards others; goal-rational, traditional and random social action are therefore operationalised to ascertain whether agents in shared environments may benefit from using alternative types of action,

in comparison to the goal-rational approaches most current systems use.

The contributions of the work in this chapter are therefore:

- The River Crossing Dilemma testbed is shown to be effective for analysing evolution and volatility when agents experience interference in shared environments; this is a valuable contribution as other testbeds in the River Crossing family only observe one agent in each environment. This testbed facilitates the exploration of the first research question outlined in Chapter 1.
- Evolving in an environment with an unknown, and unknowable other agent can potentially have a profound effect on evolution and the fitness received by agents over time; often, agents may lose their ability to achieve their goals individually. This finding relates to the first research question outlined in Chapter 1.
- Consequently, agents that share an environment with another agent experience more evolutionary volatility than when they exist alone, due to the interference arising from the actions of the other, unknown agent. This finding relates to the first research question outlined in Chapter 1, and the use of the three volatility metrics (SDoT, CACoT and CCoT) relates to the second question.
- Endowing goal-rational agents with traditional action reduces the volatility in the evolutionary process caused by interference, without affecting the fitness received. This relates to the third research question outlined in Chapter 1.
- Introducing Random Immigrants, operationalised as 'random action', is found to be quantitatively different to traditional action; this emphasises the novelty of the work in this chapter, and that it is distinct from an established way to increase diversity of populations in evolutionary algorithms. This also relates to the third research question outlined in Chapter 1.
- Further, a trade-off is observed between fitness and volatility in this study: agents that use random action experience an increase in fitness at the cost of higher evolutionary volatility, however traditional action reduces volatility without sacrificing fitness. This relates to all three questions outlined in Chapter 1.

In shared environments, the agents in these experiments experience their world changing in unanticipated and unknowable ways; they do not have the capacity to learn about or

understand the existence of others, or how their actions affect the world around them. As a consequence, this often means that agents lose their ability to achieve their goal individually, because they are unable to respond appropriately to the environmental changes caused by other agents. The focus of this study is therefore novel in comparison to other work on interference, in which knowledge of the existence of others, their goals, intentions or abilities, for example, are utilised. These results demonstrate that the type and orientation of the actions that agents take in shared environments can impact both evolutionary volatility, the fitness received, and the ability to achieve goals. Traditional action is shown to reduce the volatility experienced by goal-rational agents when they experience interference. Pursuing goal-rational action by always creating an offspring from the highest-fitness agents in a tournament (in terms of the evolutionary process) is challenged here; adding a chance for an offspring to instead be a representative state of the population (thus introducing the potential to form traditions over time) is shown to be a more effective way at reducing evolutionary volatility, and enabling agents to overcome the negative effects of interference.

As socio-technical systems grow in size and complexity, it is inevitable they will indirectly interfere with one another by interacting with the environment. Related work that studies interference utilises knowledge regarding others in the environment; in reality, such knowledge might not always be available, or may come at a high computational cost. As the field of integration science flourishes, attention is being drawn to the inherently social nature of today's systems, and both the explicit and – more importantly in this work – the *implicit* interactions in which they participate. Consequently, it is becoming increasingly necessary to enable systems to *self*-integrate with others, despite the challenges presented by an incomplete knowledge of others and the environment, uncertainty, and the task of interacting in a multitude of ways with other systems that are heterogeneous in nature. A precursor to learning about others in one's environment and how to interact with them effectively is to learn about unanticipated changes and how to manage them; this work is therefore novel by exploring how agents can mitigate interference without explicitly being aware of it.

Hähner et al. [101] argue that, as the complexity of a system increases, the number of implicit and unintended interactions will rise as a result; it would be expected then that it will become more challenging for agents to remain resilient to interference as the number of other entities in which they may potentially interact with increases. The results presented

in this chapter show the impact that social situatedness can have on an agent's ability to achieve goals in the most minimal sense – i.e. there is only one other agent to interact with. However, whilst these results provide an indication of the effect that interference can have on evolution, this may change drastically depending on the environment, task, and the number of agents within the environment; this will be a topic of future investigation.

These results suggest that endowing future systems with social self-awareness (enabling them to *intentionally* coordinate with others [28] – of which the agents in this chapter are not capable) will mean systems with many colocated components would be more capable of dealing with unanticipated and unknown interactions, due to their ability to detect and reason about others and the impact their actions can have on themselves. Further, traditional action has proven to be a beneficial step towards realising socially self-aware systems that can manage the impact of the actions of others on themselves in a socially acceptable way. While goal-rationality has been shown to be sufficient for achieving goals in a social environment, traditional action significantly reduces volatility in the evolutionary process without sacrificing fitness. Future work will explore different ways to operationalise traditional action, as well as other combinations of social action in order to observe their effects on the evolution and volatility of socially situated agents.

Chapter 5

Neuromodulation and Evolutionary Volatility

The work presented in this chapter has been adapted from the following publications:

[17] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Coevolutionary Learning of Neuromodulated Controllers for Multi-Stage and Gamified Tasks. In *Proceedings of the IEEE 1st International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 129–138. IEEE, 2020. doi: `https://doi.org/10.1109/ACSOS49614.2020.00034`.

[19] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Behavioural Plasticity Can Help Evolving Agents in Dynamic Environments But at the Cost of Volatility. *ACM Transactions on Autonomous Adaptive Systems*, 2021. In Press.

5.1 Motivation

In the previous chapter, the type of social action used by agents is shown to affect the evolutionary volatility the agents experience, and thus the ability of the agents to achieve their goals consistently over time. Specifically, traditional action (as in, acting similarly to others, where this action is oriented towards other agents in the evolutionary process) is shown to decrease volatility while not affecting the fitness that agents receive. The actions of other agents are shown to have the potential to have a dramatic effect on the ability of agents to maintain goal-achieving behaviour. However, dynamicity and the existence of interference are often characteristics observed in the natural world – in addition to the environments that artificial agents and systems are increasingly being situated in [101].

These natural and artificial environments are often complex, unpredictable and dynamic, making surviving a challenge for animals and artificial agents alike [148, 172, 242]. In order to survive in these challenging conditions, many organisms such as nematodes [223] and fish [163] show behavioural plasticity; this ability to express different behaviours – and reverse them depending on varying environmental stimuli – allows rapid adaptation to novel situations [185, 196]. This problem is not just specific to natural beings, as artificial agents and systems are often tasked with learning in dynamic and unpredictable environments too. Thus far, agents designed to operate in the RCD environment 'learn' with neuroevolution, where the behaviour that agents develop over time stems from the knowledge encoded within the weights of their deliberative neural network. However, evolving connection weights over time to encode new information can result in a degradation of performance and 'catastrophic forgetting' when learning new tasks or experiencing novel environmental contexts [146, 45, 78, 222]; learnt knowledge must be changed in order to learn new things and express new behaviours, often leading to knowledge loss [78]. Learning complex, sequential or multi-stage tasks is also difficult as complete information about the environment – including the available actions, their cues and their consequences – is not usually accessible [66, 172]. These challenges are observed in Chapter 4, since agents must learn sub-tasks to achieve their goal in the RCD, and interference from others changes the environment.

In nature, behavioural plasticity can allow organisms to adapt to novel contexts, through immediate and temporary behavioural changes depending on the stimuli experienced. One way to achieve this is with *neuromodulation* – a biological process whereby chemical signals are gated, or regulated, in the brain depending on environmental stimuli [1]. Ducatez et al. [73] demonstrate the importance of this mechanism for survival in novel environments, as species of bird capable of plastic and innovative behaviour are shown to be at a lower risk of extinction than those that cannot respond to environmental changes as effectively. As the benefit of plasticity has been widely observed in nature, researchers have taken steps to use neuromodulation to aid neural controllers with learning new or sequential tasks, and learning in dynamic environments [78, 220, 61]. Consequently, exhibiting plastic, or changeable, behaviour may help agents to mitigate interference when they are situated in shared environments, since the actions of others can make the environment appear dynamic and unpredictable.

The experimental study presented in this chapter is designed to abstract these concepts

to explore how agents evolve to solve tasks of varying complexity, when they have no knowledge of the task, environment, or others within the environment. Specifically, the notion of single- and multi-stage tasks is explored, since real-world systems may often face tasks of different complexities [48]. In this chapter, a multi-stage task is defined as one that an agent must learn and perform different behaviours in different contexts in order to achieve their goal; this definition is inspired by Dezfouli and Balleine [66]. A single-stage task is thus one that requires a singular task to be learnt. Plastic and non-plastic agents are evolved in the River Crossing Dilemma (RCD) testbed, introduced in Chapter 3, as well as an adaptation of this called the *Protected* River Crossing Dilemma (PRCD). Agents evolving in the RCD first associate the river with a negative fitness unless carrying a Stone, so multiple behaviours must be learnt depending on the agent's state; this is termed a 'multi-stage' task. The PRCD however eliminates this association between the river and negative fitness, reducing the complexity of the task to a 'single-stage' task. By investigating how agents evolve to solve different complexities of task, the effect that behavioural plasticity has on the evolution of goal-achieving behaviour in agents can be established when agents evolve in different contexts; further, the ability to temporarily change behaviour can be assessed in its suitability for mitigating interference in multi-agent environments.

Reversible and immediate behavioural changes as a result of neuromodulation are expected to enable agents to overcome the challenges associated with solving tasks and achieving goals in unpredictable and dynamic environments with greater effect than those not capable of behavioural plasticity. Neuromodulation is operationalised in this chapter by gating activation *within* a single neural network, allowing agents to regulate their behaviour without affecting encoded knowledge; this distinguishes the approach from others, which either use a separate modulatory network/neurons, or regulate learning as well as, or instead of, behaviour [220, 61, 23]. By doing this, fewer resources are required for plastic behaviour – which becomes more critical as the size or complexity of the network increases.

This study is designed to investigate the extent to which behavioural plasticity can help agents to mitigate interference in single- and multi-agent environments when others are unknown; further, how agents achieve their goals when the task and environment may vary in complexity is studied. Additionally, an analysis of the fitnesses that agents receive during and after evolution is performed, as well as the level of volatility that agents experience during evolution; these factors are important to consider when designing systems – especially

those that are in highly variable or dynamic environments – as a trade-off between fitness and predictability may need to be considered.

The remainder of this chapter is organised as follows: Section 5.2 explores the concept of behavioural plasticity in both natural and artificial environments; the design of the experimental study is explained in Section 5.3, where the PRCD testbed is introduced, and neuromodulation is operationalised; the results of the study are presented in Section 5.4 and analysed further in Section 5.5; a discussion of the results and their implications is presented in Section 5.6; finally, Section 5.7 summarises the findings and states the contributions made.

5.2 Behavioural Plasticity Arising from Neuromodulation

5.2.1 Behavioural Plasticity and Neuromodulation

One way to design adaptive systems is by utilising behavioural plasticity, which can be seen as the ability to change or adapt behaviour based on changes in environmental stimuli [148]. This is important for navigating uncertain, novel or dynamic environments and can be classed into two different types: developmental and activational [196]. Developmental behavioural plasticity can be seen as learning from experience and external stimuli. Activational behavioural plasticity on the other hand enables immediate behavioural changes; individuals can respond to new or dynamic environments during their lifetime by changing their phenotype. These behavioural changes are reversible, as the genotype remains unchanged. Activational plasticity is also termed 'innate' [148] or 'contextual' [199] plasticity.

These behavioural changes can be achieved with neuromodulation, which is a biological process found in animal brains [102]; chemical signals modify, gate or regulate synaptic plasticity based on the modulatory signal combined with the pre- and post-synaptic activities, and environmental stimuli [1, 197, 78], resulting in plastic behaviour. In neuroscience, synaptic plasticity is the modification of synapses between neurons through strengthening or weakening them [2]. In ANNs, synaptic plasticity is achieved by modulating neural network weights; short-term modifications result in immediate phenotypic changes (behaviour, appearance, etc.), and long-term changes result in learning and adaptation based on experience. Developmental plasticity is achieved by regulating learning in the long-term, where modulatory signals alter synaptic strengths; activational plasticity is achieved by regulating behaviour or synaptic activity in the short-term, without affecting learning and without

long-lasting changes to synaptic strengths.

5.2.2 Achieving Developmental Plasticity with Neuromodulation

Similarly to ANNs being inspired by the connectionist architectures found in brains, neuromodulation has been widely applied to artificial models to regulate synaptic plasticity and the learning rate of neural connections. Neural networks have been evolved with modulatory neurons to regulate learning and mitigate the catastrophic forgetting associated with performing tasks in uncertain environments [197]; this method has been found to improve learning in T-maze problems, where agents must navigate an uncertain foraging environment (either moving left or right, in a maze with a 'T'-shape) in which the location of the reward can change. Ellefsen et al. [78] however, have found that promoting the evolution of modular neural networks by introducing a cost for neural connections can mitigate catastrophic forgetting and improve learning; here, learning is regulated with neuromodulation. Neuromodulation has also been used to develop conflict learning in neural networks [99], and associative learning in robots [110]; these two approaches employ neuromodulation, but do not use neuroevolution as a learning mechanism.

The approaches outlined in this section modulate learning in the long-term by regulating the local learning rate of neurons in the network, which results in developmental plasticity; they do not however demonstrate how behaviour can be regulated in a short-term, reversible way *without* affecting learning, in order to facilitate *immediate* behavioural changes to changing environmental stimuli. Further, these approaches only use neuromodulation in neural networks or robots that exist in isolation. One of the aims of this chapter however is to explore how immediate behavioural plasticity can be achieved with neuromodulation in agents *without* regulating learning, in single- and also multi-agent environments; this may help agents to react immediately to changes within the environment as a result of interference, without necessarily changing learnt behaviour.

5.2.3 Achieving Activational Plasticity with Neuromodulation

Neurobiological mechanisms have been explored using a computational framework based on neuromodulatory systems such as the dopaminergic and serotonergic systems, by regulating synaptic activity [127]. Whilst this is proposed to aid autonomous agents in exploratory and exploitative decision-making, activational plasticity is not applied as a tool to improve

neuroevolution, but rather to explore biological systems computationally. The effects of modulating neuroreceptors and synaptic plasticity have been studied with spiking neural networks to model EEG data [79]; an aim of that work is to produce a tool to explore and diagnose neurological disorders such as dementia – and not to use neuromodulation as a tool to aid artificial agents in achieving goals. Supervised learning methods and 'context-dependent plasticity' (termed as 'activational plasticity' [196] in this chapter) have been shown to be beneficial for maintaining high accuracy for large numbers of sequential classification tasks, based on the MNIST and ImageNet datasets [143]; this was achieved by gating activations *randomly* in the network for each task. In other work, 'context-dependent selective activation' is achieved by learning parameters of a separate neuromodulatory network, which gates activity for a prediction network [23]; this two-layered neural network approach is used for learning sequential tasks by *indirectly* modulating learning, as the amount of activity in the predictive network after modulation is reflected in the back-propagation process.

Plasticity has also been explored in the context of multi-agent systems, to help agents change behaviour without changing genetic code – similar to the activational plasticity discussed in this section. Nallur and Clarke [157] introduce the concept of 'clonal plasticity', inspired by the way that plants can produce clones that are identical in terms of genetics, but can vary in phenotype depending on environmental stimuli. The authors demonstrate that clonal plasticity can lead to the emergence of diversification in agent populations in the 'Minority Game'; agents have to choose one of two options, and will receive a reward if their selection is in the minority choice selected by the population. Pitonakova et al. [168] on the other hand explore plasticity in robotic swarms, where robots are able to communicate with one another in order to forage for resources within an environment. Whilst these are interesting applications of plasticity in the domain of multi-agent systems, the focus of this chapter is instead on the modulation of neural networks to observe how agents may achieve *individual* goals when others may or may not exist within the environment. In this chapter, one agent's fitness is calculated according to its own actions; however in the work of both Nallur and Clarke [157] and Pitonakova et al. [168] (neither of which employ neural controllers), an individual's fitness is dependent on the collective actions of the population.

5.2.4 Learning Multi-Stage Tasks in Multi-Agent Environments

Both humans and animals find learning in environments that change state or context without explicit cues challenging – this however is a characteristic of most realistic environments [172]; these changes need to be detected in order to adapt behaviour accordingly, as it is rare for this information to be explicitly available. This has also been identified as being a difficulty of learning multi-stage tasks, as the full state-space of tasks is not usually available when learning [66]; despite this, agents and real-world systems often have to perform multiple tasks simultaneously [48]. These challenges are also present when neural networks learn to achieve new or many tasks, or navigate dynamic or uncertain environments; encoded knowledge must be adapted in order to learn new things, because changes in state or stimuli also change the context in which behaviours are learnt [78]. A similar problem is discussed in Chapter 4, which provides evidence for, and discusses the implications of, learning in multi-agent environments without knowledge of the existence of others; environmental stimuli change unpredictably as a result of the actions of others in the environment, which can in turn affect evolution and goal-achievement.

Regulating synaptic plasticity with neuromodulation has been shown to facilitate adaptation and learning when there are changes in environmental stimuli or the task at hand, thus helping agents to overcome these issues [61, 78, 197, 222]. Whilst neuromodulation has also been used in multi-agent contexts, this is typically to explore the effect on cooperative or competitive strategies in social dilemmas [9] or in competitive environments [242], where agents are *explicitly* aware of others and thus employ strategies intentionally. Agents acting in novel environments may not have full or even partial information about others in the environment, and thus cannot cooperate or compete intentionally. In Chapter 4, the results of the experimental study demonstrate that learning in multi-agent environments without knowledge of the existence of others is problematic due to interference. Social action is shown to improve learning in agents situated in multi-agent environments, however agents do not exhibit behavioural plasticity or use neuromodulation; furthermore, the study is limited to exploring multi-stage tasks.

The focus of this chapter is to explore the effect that behavioural plasticity arising from the regulation of activity-propagating signals has on agent evolution in varying contexts. As seen in the natural world [223], agents that are capable of expressing behavioural plasticity

would be expected to adapt better to changing, dynamic and uncertain environments, than those that are not. In artificial systems, it is common for long-term learning and activity to be regulated by a separate group of modulatory neurons or an entire network [23, 220, 61]. The study presented in this chapter has a number of distinguishing characteristics and novelties compared to other, similar work: the regulation of activity-propagating signals *within a single neural network* would promote immediate behavioural changes rather than the regulation of learning; plastic and non-plastic agent evolution is observed when agents evolve to solve either a single- or multi-stage task, in order to compare the effect of plasticity; agents are also evolved in either a single- or multi-agent environment, where other work tends to focus on single-agent evolution; additionally, neuroevolution is used in this study to evolve which neurons in a single neural network are modulatory, resulting in a more structured way of operationalising neuromodulation than in Masse et al. [143] for example, where neuronal activity is gated randomly. By not explicitly regulating learning, behaviour is instead regulated to provoke immediate phenotypic changes based on environmental stimuli. If agents are capable of changing their behaviour temporarily, they may be able to mitigate the consequences of interference with less risk of losing goal-achieving behaviour.

5.3 Experimental Study

The experiments in this study aim to investigate the effect that behavioural plasticity through activity-gating neuromodulation has on agent evolution when the environment is prone to change; the study, outlined below, is designed to explore the extent to which the ability to rapidly and reversibly change phenotypic behaviour helps agents to solve tasks in varying environmental conditions. By investigating how agents evolve to solve both learn single- and multi-stage tasks, in both single- and multi-agent environments, different combinations of environmental changes and variations are explored, and the results can be compared. Specifically, the term 'multi-stage task' is used in a similar context to Dezfouli and Balleine [66], where agents must learn multiple stages of a task in order to achieve a goal. To extend the exploration of how agents evolve in uncertain environments, the final experiments observe how agents react to a change in environmental context; to do this, agents are first evolved in an environment alone for a period of time, and then are placed within a multi-agent environment to continue evolving. Activational plasticity is hypothe-

sised to help agents to achieve their tasks in these environments, by facilitating immediate behavioural changes in response to different environmental contexts or conditions.

These concepts are important to explore since the agents in the RCD have no capacity to perceive or learn of other agents; if the environment is dynamic and uncertain due to the actions of others, the ability to immediately and reversibly change behaviour may be beneficial in an attempt to avoid failure in the face of uncertainty. In Chapter 4, traditional social action is shown to reduce evolutionary volatility compared to 'standard' goal-rational agents, without affecting fitness; at the end of the chapter, the volatility of plastic and non-plastic agents is analysed to ascertain whether a similar effect is observed.

5.3.1 Testbed Design

The River Crossing Dilemma (RCD) testbed, described in Section 3.2, is used in this study as it supports both single- and multi-agent versions of the task. It is often the case that natural and artificial agents must perform a number of actions before reaching a state capable of reaching their goals; Dezfouli and Balleine [66] use the term 'multi-stage task' to describe a task in which an agent has to perform appropriate actions in response to different environmental states in order to achieve a goal. In their study, they explore the changes in the decision-making processes of rats when they are faced with solving multi-stage tasks; this addresses an issue in other studies where subjects were given information about the task or action to perform ahead of time, as information relating to the task or state-space was not available to the rats. Here, describing the RCD task as 'multi-stage' is therefore appropriate, as agents must evolve to perform different actions depending on their own and the environment's state in order to achieve their goal; specifically, agents must learn that Water is dangerous *unless* they are carrying a Stone, and that they cannot achieve their goal unless they are able to cross the river safely – by building a bridge with Stones. In line with the study performed by Dezfouli and Balleine [66], agents in the RCD testbed have no *a priori* knowledge of the task or the environment, and must learn the appropriate behaviours to respond to different states and environmental stimuli.

The Protected River Crossing Dilemma

In addition to the RCD which incorporates a multi-stage task, here, the Protected River Crossing Dilemma (PRCD) is introduced; this is an adaptation of the RCD specifically

designed to explore how agents evolve to solve single-stage tasks, since actors in both natural and artificial environments may face tasks of different complexities [48].

The environment is constructed identically to the RCD as described in Section 3.2 (including gamification of the testbed, and the fitness function used), with the sole difference that the river acts as an impassable – and most importantly, a non-lethal – obstacle; this means that agents cannot move into the river and consequently drown, and can only interact with Water by placing a Stone to build a bridge. This simple change removes the need for agents to learn the different states in which they can interact with the river, i.e. that it is not safe unless the agent is carrying a Stone. As the PRCD river is impassable, agents must still perform sub-tasks such as bridge-building to achieve their goal of collecting their reward object from the opposite river bank; removing the river entirely would remove the multi-stage task but also make the task trivial. This single-stage version of the RCD thus reduces the variability in the task and environment, making it less complex; as the degree of plasticity is said to increase with environmental variability [122], the effect of neuromodulation is expected to be less apparent in single-agent environments of both the RCD and PRCD, than in multi-agent versions of the tasks.

5.3.2 Agent Design

The agents in both the RCD and the PRCD environments use the two-tiered neural network architecture described in Section 3.3. Specifically, the deliberative network, which can be seen in Figure 5.1, has three hidden layers with eight, six and four neurons respectively; all other details of the agent design are as previously specified. The version of the deliberative neural network used in this chapter has an increased number of neurons compared to the network used in Chapter 4 (Figure 3.2); this is intended to simulate a larger brain size than the previous study. Snell-Rood [196] posits that activational behavioural plasticity – the focus of this chapter – increases with brain size, in terms of the number of neurons; Herczeg et al. [103] observe this effect in guppies, where brain size can indicate the degree of behavioural plasticity and an individual's ability to adapt to novel environments. Increasing the number of neurons in the deliberative neural network in this study compared to the previous study is therefore intended to give agents the chance to express a higher degree of plastic behaviour than would be expected if a smaller network was used.

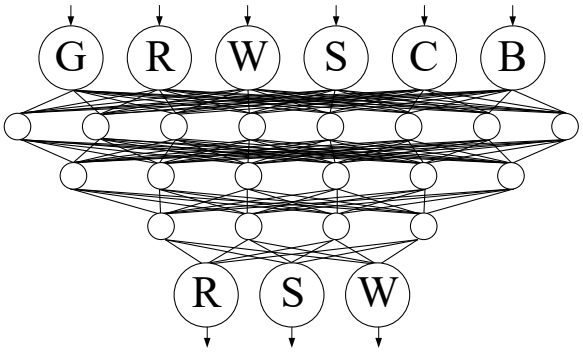


Figure 5.1: The deliberative network is a feed-forward neural network that generates high-level sub-goals – here, with three hidden layers. Inputs are 1 or 0, corresponding to the agent's current state: Grass (G), Resource (R), Water (W), Stone (S), Carrying Status (C), if a Bridge partially exists (B). Outputs are 1 for attraction, 0 for neutral or -1 for avoidance for each sub-goal: Resource (R), Stone (S), Water (W).

5.3.3 Operationalising Activity-Gating Neuromodulation

In this study, neuromodulation is operationalised as follows. Modulatory agents (also referred to as 'plastic') regulate their behaviour by gating and temporarily suppressing activity *within* the deliberative network (Figure 5.1), without permanently changing the weights of the network; this distinguishes the approach proposed in this chapter from others, which either use a separate modulatory network/neurons, or regulate learning as well as, or instead of, behaviour [220, 61, 23]. As behavioural plasticity is said to increase survival in novel and dynamic environments (Section 5.2), it would be expected that this type of immediate, temporary and reversible plasticity, achievable with activity-gating neuromodulation, would enable agents to achieve their goals with a higher chance of success in the RCD. Further, it would be expected that plasticity would also be beneficial to agents in shared environments, as agents have no ability to perceive others, and therefore perceive the actions of others as dynamicity in the environment itself.

Figure 5.2 shows an example of this activity-gating modulation. In this approach, there are two different types of neuron in the deliberative neural network – a standard, non-modulatory neuron, and a modulatory neuron; the hidden neurons in the deliberative neural network may evolve to be non-modulatory or modulatory (explained further in Section 5.3.4). The output of a 'standard' non-modulatory neuron is calculated in the same way as defined in Section 3.3, by summing each input signal multiplied by the weight of the connection, and passing this through an activation function (*tanh* in this case); the output activation is then passed along the outgoing connections. Modulatory neurons also propagate activity in the same way if the incoming signal (sum of inputs) is positive. If the

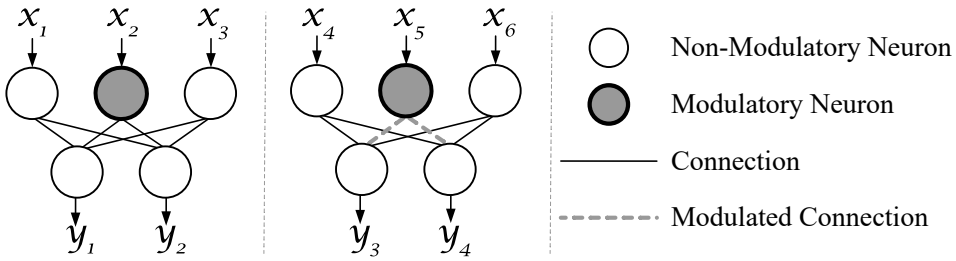


Figure 5.2: In activity-gating neuromodulation, modulatory neurons propagate activity the same as non-modulatory neurons when the incoming signal is ≥ 0; here, if the incoming activity signal to x_2 is positive, the outgoing activity signals of x_2 propagate as usual and are passed on to the next layer of neurons (in this case, y_1 and y_2). If, however, the incoming activity signal to x_5 is negative, the modulatory neuron fires and the outgoing activity is gated; specifically, this means that the neuron x_5 will output signals of 0 along each of its outgoing connections (in this case to y_3 and y_4), so the outgoing signal is effectively gated or 'turned off' when the signal is multiplied by the weight of the connection. This means agents can exhibit behavioural plasticity, as the weights of the neural network are not *changed*, but temporarily suppressed; this leads to the network producing different outputs and therefore different behaviours, without modifying the network weights in a permanent way. It is important to note that modulatory neurons only affect their own outgoing connections, so the connections from x_4 and x_6 to y_3 and y_4 are not affected when x_5 fires.

incoming signal to a modulatory neuron is negative, the neuron will regulate the agent's behaviour by outputting a signal of 0 on each of its outgoing connections. This means that weights on the connections will effectively be 'turned-off', or gated, as the signal is blocked locally. Immediate, and more importantly reversible behavioural changes can therefore be achieved depending on the stimuli experienced; the weights of the connections are therefore not modified in a permanent way, which would instead regulate learning in the long-term rather than *behaviour* in the short-term.

This gating or *modulation* of activity-propagating signals results in behavioural plasticity; an agent's genotype, represented by the evolved weights of the neural network and the types of the neurons in the deliberative network, is therefore able to express multiple phenotypes depending on state and environmental stimuli – without changing, or potentially destroying, the knowledge encoded in the weights of the network. In other words, a modulatory agent can temporarily change behaviour depending on the stimuli and inputs; this is because modulatory neurons that are 'switched off' do not propagate any activity signals to the next layer of neurons, thus changing the output of the network and the resulting behaviour of the agent.

5.3.4 Neuromodulation in the Evolutionary Algorithm

The agents in this study are evolved using the Steady State Genetic Algorithm specified in Section 3.4. For modulatory agents, the neurons in the hidden layers of the delibera-

tive network are evolved in addition to the weights (input and output neurons cannot be modulatory); neurons may evolve to be standard non-modulatory neurons, or modulatory neurons that regulate network activity. The deliberative network of each agent is initialised with only non-modulatory neurons, then the neuron types are evolved with neuroevolution like the weights of the network. Agents that do not use neuromodulation have a static network of non-modulatory neurons that do not evolve.

At each generation, the new offspring inherits the neuronal structure from a randomly chosen parent, where the parents are the two agents with the best fitnesses in the tournament; all other details of the Steady State Genetic Algorithm, including how the weights of the network are mutated, are as specified in Section 3.4. There is an additional probability of $P_{mut} = 0.15$ that one randomly chosen hidden neuron in the deliberative network (Figure 5.1) will be mutated, from non-modulatory to modulatory or vice versa. This mutation rate is inspired by the mutation operators and probabilities used by Ellefsen et al. [78], where neuromodulation is used to regulate learning when evolving modular neural networks to reduce catastrophic forgetting; the aim of this work has similarities to that of this study, as neuromodulation here is used to help agents learn multiple sub-tasks simultaneously in dynamic environments.

5.3.5 Experimental Design

All experiments are repeated 100 times using the same 100 seeds, both with and without neuromodulation. The first four sets of experiments evolve agents for 500,000 generations from a randomly-initialised state; the final two sets of experiments evolve agents for 1,000,000 generations in total by first evolving agents in an environment alone, and then continuing to evolve them in a shared environment together with another agent.

The first set of experiments explore how agents evolve to solve a single-stage task in the PRCD, when they exist alone in the environment. This environment has the least inherent variability, which will provide a baseline to compare the effects of neuromodulation in the later experiments. Variability increases if there is more than one agent in the environment, since the actions of each agent can change the environment unpredictably.

The second set of experiments introduces another agent into the PRCD environment; this creates a social dilemma, so agents may evolve to cooperate or exploit the other unintentionally. As agents cannot perceive or reason about the actions or existence of other

agents, their environment appears unpredictable and is harder to evolve in than when existing alone. These experiments evolve two separate, randomly-initialised populations of agents that start on opposite corners of the environment.

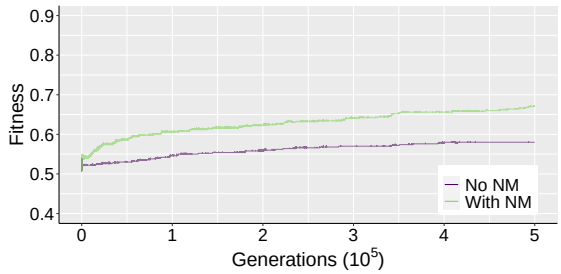
The third set of experiments explores environmental variability further, by evolving agents alone for an initial period of 500,000 generations, and observing how they are able to achieve their goals when they continue to evolve in a shared PRCD environment with another agent for a further 500,000 generations. Agents therefore evolve for a total of 1,000,000 generations, with the initial period of evolving alone being identical to the first set of experiments in this study. These experiments are thus termed as 'continued evolution' (CE), as agents continue to evolve with another after evolving in isolation. Here, the extent to which activity-gating neuromodulation affects how agents achieve goals when the environment explicitly changes context from a single- to a multi-agent environment.

The fourth, fifth, and sixth sets of experiments are repetitions of the first three sets, however these are conducted in the RCD environment. This multi-stage version of the task requires agents to learn more behaviours than in the PRCD, since agents are able to interact with or drown in the river; as such, it would be expected that this task is harder for agents to learn due to the risk associated with the river.

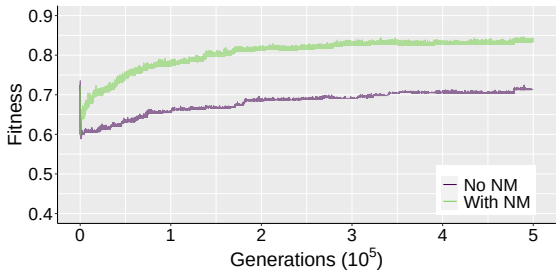
5.4 Results

5.4.1 Learning Single-Stage Tasks When Alone

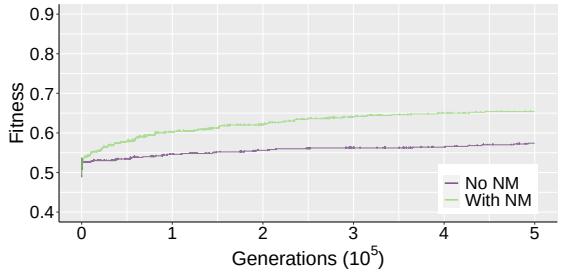
Firstly, agents are observed when evolving to solve the simplest task in the least variable environment in the study – the single-stage task in the PRCD; the effect of neuromodulation on agent evolution can be ascertained by comparing the evolution of both plastic and non-plastic agents in these experiments. Figure 5.3a shows the mean best-in-population fitness over time when agents evolve alone in the PRCD, both with and without neuromodulation. The benefit of neuromodulation is seen from the beginning of, and is sustained throughout, evolution. 85% of modulatory agents were able to solve the single-stage task, compared to only 40% of agents that did not use neuromodulation (Table 5.1).



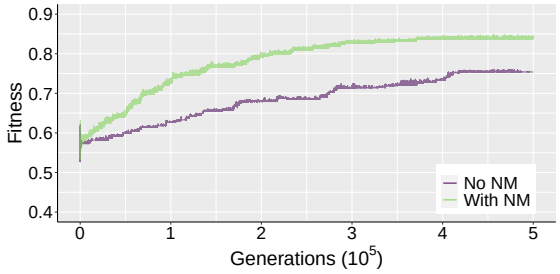
(a) Single-Stage Task – Alone



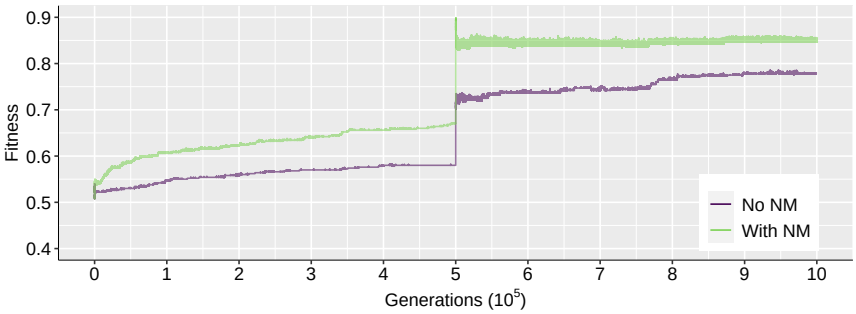
(b) Single-Stage Task – Together



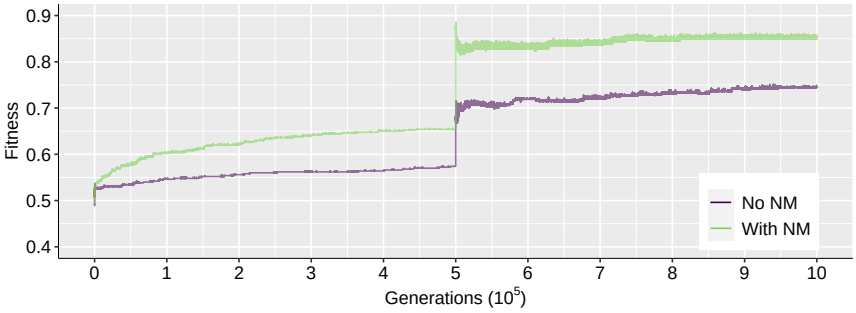
(c) Multi-Stage Task – Alone



(d) Multi-Stage Task – Together



(e) Single-Stage Task – Continued Evolution



(f) Multi-Stage Task – Continued Evolution

Figure 5.3: The mean best-in-population fitnesses of agents evolving to solve (a) a single-stage alone or (b) together, (c) a multi-stage task alone or (d) together, and (e) a single-stage or (f) a multi-stage task with continued evolution, with and without neuromodulation (NM). Single-stage and multi-stage tasks take place in the PRCD and RCD respectively. A fitness of: 0.7 indicates the goal is achieved individually; 0.9 indicates the cost of bridge-building is shared; 1.0 indicates exploitation; 0.7 or above indicates the goal is achieved; below 0.7 indicates the task is failed (Equation 3.1).

5.4.2 Learning Single-Stage Tasks When Together

By introducing two agents into the single-stage task PRCD environment, the variability of the environment increases, and the task becomes gamified. To achieve their goal, agents may evolve to achieve their goal alone, cooperate unintentionally, or exploit the actions of the other agent; agents therefore have the potential to achieve a higher fitness, but at the risk of becoming dependent on the actions of another to achieve their goal. Agents are unable to perceive others or their actions, so the environment becomes unpredictable when it is shared with another agent; the resulting uncertainty is perceived to be a characteristic of the environment, rather than arising from another agent.

Figure 5.3b shows the mean best-in-population fitness of agents evolving together in a shared PRCD environment. Similarly to when agents evolve to solve a single-stage task alone (Figure 5.3a), neuromodulation is beneficial from the start of evolution. In Figure 5.3b, modulatory agents evolve to achieve a higher fitness more often, and by the end of evolution, 97% of modulatory agents achieve their goal compared to 61% of non-modulatory agents (Table 5.1). The effect of neuromodulation is more prominent when agents evolve together compared to when they evolve alone, as agents can achieve a higher fitness. This finding is not interesting in itself, however the fact that fewer agents evolve to achieve their goals individually in shared environments compared to those in isolation (Table 5.1) demonstrates the impact that interference can have on the evolution of goal-achieving behaviour. Relying on other agents to achieve goals can be detrimental if those agents change their behaviour or leave the environment. Further, the spike in fitness at the beginning of evolution is caused by both agents reacting to and evolving based on the changes in the other's behaviour; once each agent's behaviour becomes more predictable, this spike drops. This is also observed in Figure 5.4.5.

5.4.3 Learning Single-Stage Tasks with Continued Evolution

When another agent is introduced into the PRCD environment after an agent has evolved in isolation for a period of time, the perceived variability of the environment increases; an environment that was once predictable then becomes unpredictable, due to the actions of the new agent. By continuing to evolve in this new, shared environment, behaviour that was reliable in isolation may no longer be reliable if the environment changes in unknown

Experiment	Task (S/M)	NM	Fitness (% of Agents)				
			0.7	0.9	1.0	< 0.7	≥ 0.7
Alone	S	No	40	0	0	60	40
		Yes	85	0	0	15	85
	M	No	37	0	0	63	37
		Yes	77	0	0	23	77
Together	S	No	29	5	27	39	61
		Yes	49	2	46	3	97
	M	No	27	5	36	32	68
		Yes	44	0	50	6	94
Continued Evolution	S	No	34	10	34	22	78
		Yes	46	11	43	0	100
	M	No	40	1	32	27	73
		Yes	47	2	50	1	99

Table 5.1: The percentage of agents that receive common fitnesses in each experiment, after 500,000 generations of solving a single- (S) or multi- (M) stage task, with or without neuromodulation (NM). 0.7 is a goal-achieving fitness after a bridge is built with two Stones; 0.9 is sharing the cost of bridge-building; 1.0 is exploitation; < 0.7 does not achieve the goal; ≥ 0.7 is a goal-achieving fitness.

ways. Not only this, but the opportunity to receive a higher fitness becomes accessible when there is another agent introduced into the environment, as one agent may cooperate with or exploit the other.

The mean best-in-population fitness of agents evolving with continued evolution in the PRCD is shown in Figure 5.3e. There is a dramatic increase in fitness in the final 500,000 generations when the agents continue to evolve with another agent in the environment. Modulatory agents appear to benefit more from this change in variability as 100% of modulatory agents achieve their goal after continued evolution, compared to only 78% of non-modulatory agents (Table 5.1).

5.4.4 Learning Multi-Stage Tasks When Alone

Evolving to solve the multi-stage task present in the RCD is a more challenging – and more perilous – task when the possibility of falling in the river exists, compared to the single-stage task in the PRCD; agents must evolve to match correct behaviours with different environmental stimuli under different conditions in order to succeed. When agents evolve alone in the RCD environment, they can only achieve their goal once they have built a bridge on their own. The maximum fitness an agent can achieve is therefore 0.7, after the bridge-building cost is deducted from the fitness (Equation 3.1).

The mean best-in-population fitness increases over time as more agents evolve successful solutions; after 500,000 generations, 37% of agents evolved the necessary behaviours to achieve their goal without neuromodulation, compared to 77% that achieved their goal with neuromodulation (Table 5.1). Figure 5.3c shows that the mean best-in-population fitness is higher when agents use neuromodulation, indicating that agents are more likely to evolve successful solutions, and that they are able to do this in fewer generations than agents that do not use neuromodulation.

5.4.5 Learning Multi-Stage Tasks When Together

The fitness function presented in Equation 3.1 evaluates each agent individually. When agents share an environment, they can still achieve their goal alone by building a bridge completely by themselves and enduring the associated cost; they can also exploit the other to avoid the cost, or cooperate to share the cost of bridge-building. The maximum fitness accessible to each agent therefore increases to 1.0 instead of 0.7, as agents may achieve their goal without building a bridge. In each case, agents have no capacity to perceive the existence or actions of the other, so cannot cooperate or exploit intentionally. The multi-stage task in the RCD adds yet another layer of complexity onto the task and the environment; a multi-agent environment introduces an element of unpredictability as agents cannot perceive others, and a multi-stage task means that the agent must discover multiple states and the corresponding consequences in the environment in order to achieve the goal.

Figure 5.3d shows that modulatory agents evolve to achieve their goal more often, and in fewer generations, than non-modulatory agents. After 500,000 generations, 94% of modulatory agents achieve their goal, compared to only 68% of non-modulatory agents (Table 5.1). This shows that behavioural plasticity is beneficial for agents to respond to changes in environmental stimuli caused by the actions of others. Agents in multi-agent environments are affected by the actions of the other agent in one way or another; this is seen when agents solve single- and multi-stage tasks. Table 5.1 shows that fewer agents achieve their goals individually (by building a bridge on their own, to receive a fitness of 0.7) when evolving together, than when evolving alone; overall, more agents achieve their goals in shared environments because some exploit or cooperate with the other agent, but this may be detrimental in the long run if agents are unable to learn bridge-building behaviour themselves. Additionally, agents that learn a multi-stage task achieve their goals individually less

often than those learning a single-stage task, when comparing each experiment presented in Table 5.1. This gives evidence that learning a multi-stage task can be more difficult for agents to learn than a single-stage task, and that evolving goal-achieving behaviour is thus more challenging when the complexity of the task is increased. This observation is less prominent in multi-agent environments, since agents may simply learn to exploit others to achieve their goal, instead of evolving the behaviours to achieve it individually.

5.4.6 Learning Multi-Stage Tasks with Continued Evolution

The maximum fitness agents can achieve in the RCD rises from 0.7 to 1.0 when agents change from being in isolation to a shared environment. In the following experiments, agents evolve alone in the multi-stage RCD environment for 500,000 generations, then continue to evolve with another for a further 500,000 generations. Agents must adapt their behaviour to cope with environmental changes, and the unanticipated actions of others, since the context in which the agents have evolved in is completely changed.

Figure 5.3f shows the mean best-in-population fitness of agents that continue to evolve together. The change in context from a single-agent to a multi-agent environment has an instantaneous effect on the evolution of agents, which can be seen in the spike in fitness at generation 500,000. This is because both plastic and non-plastic agents can immediately capitalise on the changes in environmental stimuli caused by the imperceptible actions of the other agent in the environment. This spike in fitness then falls slightly while agents adjust to the new change in context. Agents are evolving in tandem and changing their behaviour in response to the other agent's changes in behaviour; the fitness falls slightly after the change in context occurs as agents learn that other agents might not always be reliable, and thus evolve to achieve lower fitness by achieving goals alone. This is also seen in Figure 5.3e. Neuromodulation is observed to help agents to adapt to their new, shared environment when the context of the task is changed; 99% of modulatory agents achieve their goal, compared to only 73% of non-modulatory agents. Similarly to the other experiments, modulatory agents achieve a higher fitness earlier in evolution than their non-modulatory counterparts, which indicates that behavioural plasticity can increase the speed in which goal-achieving solutions are found, and consequently maintained during evolution.

Generally, fewer agents are able to achieve their goal individually with continued evolution in these experiments, compared to when they are alone. More agents are able to

individually achieve their goal in the multi-stage task environment than the single-stage task, although fewer agents achieve their goal overall. The effect that changing the context in which an agent evolves affects evolution is studied in more detail in Chapter 6.

5.5 Analysing the Effect of Plasticity on Agent Evolution

Activity-gating neuromodulation increases both the likelihood and the speed that agents evolve successful solutions – both when they exist alone, and when they exist together (Figure 5.3). This section analyses agent evolution in each experiment further, and aims to ascertain whether behavioural plasticity affects agents in terms of the fitness they receive both at the end of and during evolution. Additionally, the extent to which neuromodulation affects the volatility that agents experience in terms of fluctuations in fitness during evolution is explored.

5.5.1 Analysing Fitness

Analysing the Fitness Distributions

The statistical moments and median for the best-in-population fitness of each experiment were calculated at the end of evolution, presented in Table 5.2. This analysis shows that modulatory agents have the same or higher mean and median fitness across all experiments. Combined with the results presented in Table 5.1, modulatory agents therefore not only have a higher mean and median fitness, but they achieve their goal more often than non-modulatory agents in each experiment. The variance in the best-in-population fitness after evolution is also less in modulatory agents, which further illustrates the benefits of behavioural plasticity; agents will predictably receive higher fitnesses with neuromodulation.

The distribution of fitnesses after evolution for modulatory agents is negatively skewed, where the amount of skewness tends to decrease from being highly skewed to more symmetrical as environmental variability increases. An increase in environmental variability here refers to the way the environment changes unpredictably due to the actions of other agents; environments where agents exist alone thus have less variability. This can be seen in Figure 5.4, as the peaks in the kernel density estimation plots are towards the higher end of the x-axis. This is supported by the fact that the median fitness tends to be higher than the mean fitness for modulatory agents, meaning that agents would be likely to achieve a

Exp	Task	NM	Moment				Median
			Mean	Variance	Skewness	Kurtosis	
Alone	S	No	0.580	0.009 70	0.408 *	1.17*	0.500
		Yes	0.670*	0.005 15*	−1.96	4.84	0.700*
	M	No	0.574	0.009 42	0.539 *	1.29*	0.500
		Yes	0.654*	0.007 16*	−1.28	2.65	0.700*
Together	S	No	0.713	0.0422	0.345	1.55	0.700
		Yes	0.836*	0.0252 *	−0.0856*	1.42*	0.700
	M	No	0.754	0.0447	0.0245*	1.36*	0.700
		Yes	0.838*	0.0286 *	−0.282	1.60	0.850*
Continued Evolution	S	No	0.778	0.0377	−0.137	1.56	0.700
		Yes	0.851*	0.0205 *	−0.0513*	1.10*	0.900*
	M	No	0.744	0.0384	0.195	1.61	0.700
		Yes	0.852*	0.0233 *	−0.132 *	1.21*	0.950*

Table 5.2: Statistical moments and median (to 3 S.F.) of the best-in-population fitness after 500,000 generations of evolving alone or together, or after 1,000,000 generations of continued evolution. Agents evolve to solve a single- (S) or multi- (M) stage task. The highest mean and median, and lowest variance, skewness, and kurtosis for agents with or without neuromodulation (NM) are indicated with an asterisk (*).

higher-than-average fitness. The opposite is generally true in non-modulatory agents, as the fitness distributions are positively skewed; as with the modulatory agents, the amount of skew tends to decrease as environmental variability increases. In each experiment, the mean fitness for non-modulatory agents is higher than the median; this indicates that the distributions are positively skewed and that agents would be likely to achieve a fitness less than the average. A contributing factor to this is that non-modulatory agents are less likely to evolve a goal-achieving fitness at the end of evolution than modulatory agents; this can be seen in Figure 5.4, as the distribution is skewed to the right. The exception to this is when non-modulatory agents evolve to solve a single-stage task with continued evolution, as the distribution is negatively skewed; this is influenced by a large number of agents achieving high fitnesses at the end of evolution in this experiment (78%, Table 5.1) compared to non-modulatory agents in the other experiments.

The kurtosis in the fitness distributions tends to increase in non-modulatory agents as environmental variability increases, but decrease in modulatory agents; this suggests that more outliers can be expected in the fitness distributions of non-modulatory agents as environmental variability increases, and the opposite in modulatory agents. Saying this, all fitness distributions are platykurtic (where excess kurtosis is negative ($kurt_{excess} = kurt - 3$), or $kurt < 3$), meaning that outliers and extreme values are not common overall.

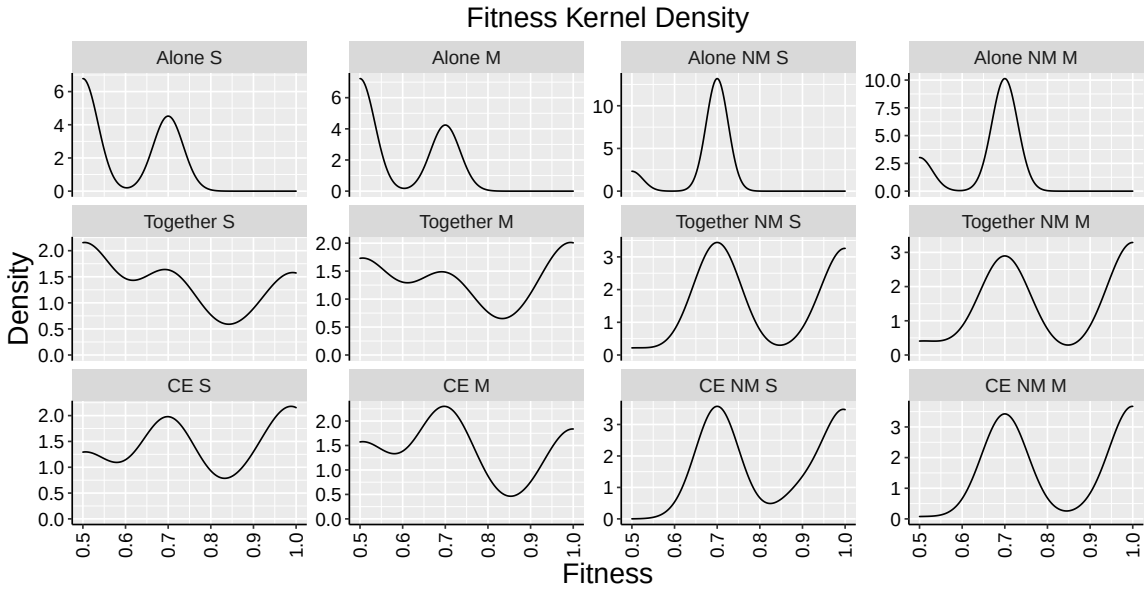


Figure 5.4: Kernel density estimations of the fitness of agents that evolve alone, together, or with continued evolution (CE), and with or without neuromodulation (NM), to solve a single- (S) or multi- (M) stage task. *Note: graph scales are not comparable due to the disparity in the densities of each distribution.*

Statistical Tests – Comparing Non-Modulatory and Modulatory Agents

To analyse the effect that activity-gating neuromodulation has on evolution further, statistical tests were performed to compare the best-in-population fitnesses of modulatory and non-modulatory agents in each experiment. Firstly, a Shapiro-Wilk test is described by Yap and Sim [241] as being a powerful test for normality for a range of distributions that are skewed, symmetric, and those with high or low kurtosis; as such, this is appropriate to test the distributions described by the statistical moments and median in Table 5.2. Each distribution was found to be non-normal, with the p-value for each Shapiro-Wilk test being below 0.05; the results of these tests can be found in Table B.4 in Appendix B.2.

As the distributions are non-normal, Wilcoxon Signed Rank statistical tests were then conducted to analyse the effects of behavioural plasticity on fitness and evolution; here, the non-modulatory (m_n) and modulatory (m_m) approaches are compared, and the results are presented in Table 5.3. The two-tailed tests show that there is a significant difference in median fitness between non-modulatory and modulatory agents, for each experiment in the study ($p < 0.05$, $m_n \neq m_m$); the null hypothesis that the medians of the two distributions are equal, can thus be rejected as $p < 0.05$. Further, the median fitness of non-modulatory agents (m_n) is found to be significantly lower than in modulatory agents

Exp	Task (S/M)	Statistical Test Alternative Hypothesis			z	r
		$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$		
Alone	S	2.588×10^{-9} *	1.294×10^{-9} *	1	−5.960	−0.5960 (L)
Together	S	2.362×10^{-4} *	1.181×10^{-4} *	9.999×10^{-1}	−3.679	−0.3679 (M)
CE	S	1.003×10^{-2} *	5.014×10^{-3} *	9.951×10^{-1}	−2.580	−0.2580 (S)
Alone	M	2.994×10^{-8} *	1.497×10^{-8} *	1	−5.547	−0.5547 (L)
Together	M	1.594×10^{-2} *	7.970×10^{-3} *	9.922×10^{-1}	−2.413	−0.2413 (S)
CE	M	2.593×10^{-4} *	1.296×10^{-4} *	9.999×10^{-1}	−3.656	−0.3656 (M)

Table 5.3: Wilcoxon Signed Rank statistical tests comparing the best-in-population fitness at the end of evolution of the non-modulatory (m_n) and modulatory (m_m) agents in each experiment: evolving in an environment alone, together, or with continued evolution (CE), with a single- (S) or multi- (M) stage task. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

(m_m) for each experiment conducted ($p < 0.05$, $m_n < m_m$). These results demonstrate that neuromodulation is beneficial to agents by increasing the observed fitness in all areas of the study, however, the strength of this benefit is as of yet unexplored.

An effect size estimate r can be used to ascertain the strength of a relationship between two variables; here, the best-in-population fitnesses after evolution of non-modulatory and modulatory agents in each experiment are analysed. A negative effect size estimate r in each of these experiments, seen in Table 5.3, indicates that modulatory agents receive a higher fitness than non-modulatory agents. Further, it appears as though behavioural plasticity via neuromodulation has a stronger effect on agents that are situated in an environment alone, compared to those that act in a shared environment; one contributing factor to this could be that agents can access a higher fitness when in a shared environment by exploiting the actions of another agent – thus increasing the similarity between the fitnesses that non-modulatory and modulatory agents receive overall. One thing to note here is that although an agent can achieve higher fitnesses by cooperating with or exploiting another (thus increasing the similarity between modulatory and non-modulatory agents as stated above), this does not necessarily mean that the agent will have evolved the necessary behaviours to achieve its goal alone; whilst achieving a high fitness by any means is desirable, it is important for agents to be capable of achieving the goal if they find themselves unable to depend on the other. Additionally, behavioural plasticity has a 'medium' effect [58] on the fitness agents receive when they evolve alone for an initial period of time, and then continue to evolve in a shared environment ('continued evolution'). Whilst effect size is not linear, neuromodulation has a stronger effect on the fitness of agents evolving with continued evo-

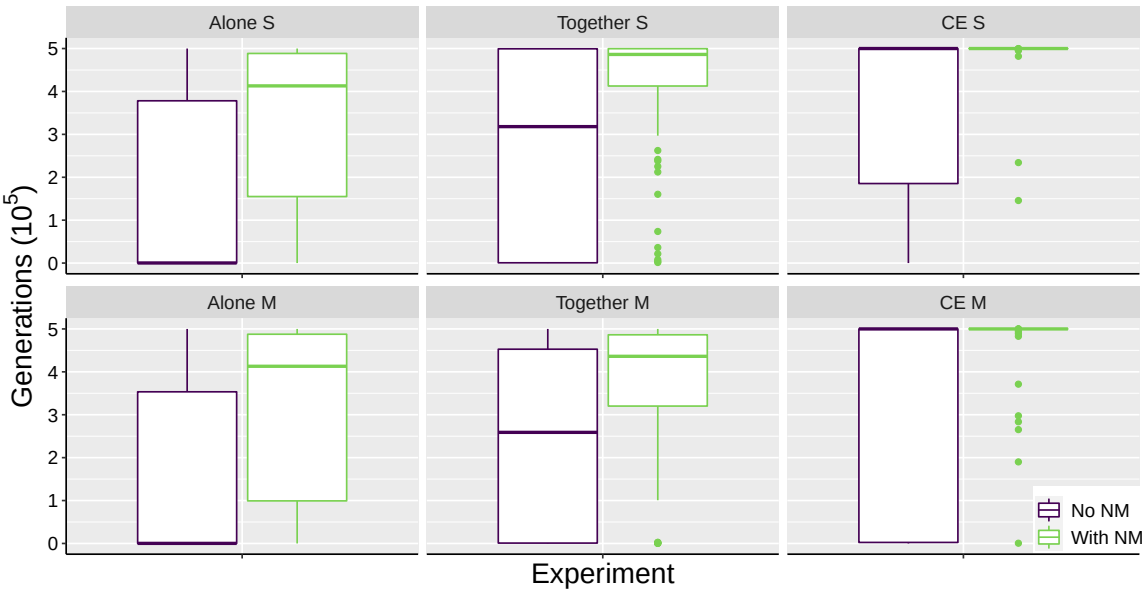


Figure 5.5: Box plot depicting the number of generations agents receive a goal-achieving fitness (≥ 0.7) during 500,000 generations of evolving to solve a single- (S) or multi- (M) stage task; agents evolve alone, together, or with continued evolution (CE, excluding the initial period of evolving alone), and with or without neuromodulation.

lution than on those that evolve together to solve a multi-stage task, but a weaker effect than on those that evolve alone; this is an unsurprising observation as 'continued evolution' here encompasses evolution in both single- and multi-agent environments.

5.5.2 Analysing Goal-Achievement over Evolution

Thus far, the fitness that agents receive at the end of evolution has been assessed; modulatory agents are observed to have a higher mean fitness than non-modulatory agents, and are also more likely to achieve their goals. However, while it is desirable to evolve agents that can receive a goal-achieving fitness at the end of evolution, another benefit would be for agents to consistently achieve their goals throughout evolution as well.

Figure 5.5 shows a box plot of the number of generations that agents in each experiment receive a goal-achieving fitness (≥ 0.7) during 500,000 generations of evolution; the detailed statistics can be found in Table 5.4. In each experiment, the first, second and third quartiles are the same or higher in modulatory agents than in non-modulatory agents; this shows that modulatory agents achieve their goal for more generations overall than their non-modulatory counterparts. Not only is the data more heavily skewed to the left in modulatory agents, but the spread of values is generally smaller than in non-modulatory agents; this indicates that

Experiment	Task	NM	Min	Q_1	Median	Mean	Q_3	Max
Alone	S	No	0	26	217	151 588	378 222	500 000
		Yes	0	155 103	412 986	316 252	488 567	500 000
	M	No	0	1	170	139 129	353 437	500 000
		Yes	0	99 206	412 955	310 455	487 648	500 000
Together	S	No	46	653	318 043	251 308	499 397	499 968
		Yes	1394	412 690	486 264	423 809	499 654	499 980
	M	No	16	852	258 916	244 726	452 762	499 992
		Yes	399	320 212	436 195	385 511	486 331	499 977
Continued Evolution	S	No	0	185 382	500 000	354 842	500 000	500 000
		Yes	145 809	500 000	500 000	493 500	500 000	500 000
	M	No	0	2383	500 000	336 102	500 000	500 000
		Yes	615	500 000	500 000	483 722	500 000	500 000

Table 5.4: The minimum and maximum number of generations that agents achieve their goal, along with the first, second and third quartiles (Q_1, median, and Q_3 respectively), and the mean. Agents evolve for 500,000 generations to solve a single- (S) or multi- (M) stage task, when alone, together, or with continued evolution (excluding the initial period of evolving alone), and with or without neuromodulation (NM). Values are rounded to the nearest whole number.

modulatory agents are more predictable, and are likely to spend more generations receiving a goal-achieving fitness than other agents. Agents using neuromodulation thus spend more of evolution able to achieve their goals than agents not capable of behavioural plasticity.

To evidence this claim further, Wilcoxon Signed Rank statistical tests were conducted to compare the number of successful generations between non-modulatory and modulatory agents, where a 'successful' generation is one that an agent receives a goal-achieving fitness of ≥ 0.7 (Table 5.5). In line with Section 5.5.1, a Shapiro-Wilk normality test first indicated each distribution was non-normal ($p < 0.05$); the results can be found in Table B.5 in Appendix B.2. This justifies the use of a Wilcoxon Signed Rank statistical test. In each experiment, the median number of successful generations is found to be significantly different between non-modulatory and modulatory agents ($p < 0.05$, $m_n \neq m_m$). Further, there is a significant directional difference between the two medians; the median number of generations that non-modulatory agents successfully achieve their goals is less than in modulatory agents ($p < 0.05$, $m_n < m_m$). The analysis thus far shows that behavioural plasticity is beneficial to agents solving a variety of tasks under different environmental conditions, as modulatory agents are more likely to receive a higher fitness at the end of evolution, achieve their goals, and be more successful throughout evolution than non-modulatory agents. Further evidence to support this conclusion is that a negative effect size estimate r, seen in Table 5.5, is observed in each experiment; this indicates that modulatory agents are successful for more

Experiment	Task (S/M)	Statistical Test Alternative Hypothesis			z	r
		$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$		
Alone	S	2.128×10^{-7} *	1.064×10^{-7} *	1	−5.190	−0.5190 (L)
Together	S	1.846×10^{-7} *	9.230×10^{-8} *	1	−5.216	−0.5216 (L)
Continued Evolution	S	6.973×10^{-9} *	3.487×10^{-9} *	1	−5.800	−0.5800 (L)
Alone	M	3.031×10^{-7} *	1.515×10^{-7} *	1	−5.123	−0.5123 (L)
Together	M	2.177×10^{-6} *	1.088×10^{-6} *	1	−4.738	−0.4738 (M)
Continued Evolution	M	2.484×10^{-8} *	1.242×10^{-8} *	1	−5.578	−0.5578 (L)

Table 5.5: Wilcoxon Signed Rank statistical tests comparing the number of generations that the best-in-population non-modulatory (m_n) and modulatory (m_m) agents receive a goal-achieving fitness (≥ 0.7) in each experiment: evolving alone, together, or with continued evolution, with a single- (S) or multi- (M) stage task. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

generations during evolution than non-modulatory agents, regardless of whether the agents evolve to solve a single- or multi-stage task, in a single- or multi-agent environment.

5.5.3 Analysing the Effect of Behavioural Plasticity on Volatility

The results of the experimental study show that behavioural plasticity arising through neuromodulation has a positive impact on the fitness agents achieve after evolution, as well as the ability of agents to achieve their goals in both single- and multi-agent environments. In this section, the impact that behavioural plasticity has on the evolutionary volatility of agents – and thus their evolved fitness over time – is investigated. The three volatility metrics outlined in Section 3.6.1 are used as a means of analysing the evolutionary volatility that agents experience in each experiment; these metrics can be used to describe the evolutionary process of agents, and whether the received fitness is prone to change frequently during evolution. For all 100 runs of each experiment, a value for each of the three metrics was calculated using the best-in-population fitness at each generation across 500,000 generations of evolution, or all 1,000,000 generations for agents evolving with Continued Evolution. Statistical moments and medians are presented for each metric in Table 5.6, and kernel density estimation plots for the SDoT, CACoT and CCoT distributions are presented in Figures 5.6, 5.7 and 5.8 respectively.

Analysing Volatility

In all experiments, non-modulatory agents are observed to have a lower mean and median SDoT, CACoT and CCoT (Table 5.6) than their modulatory counterparts; this indicates

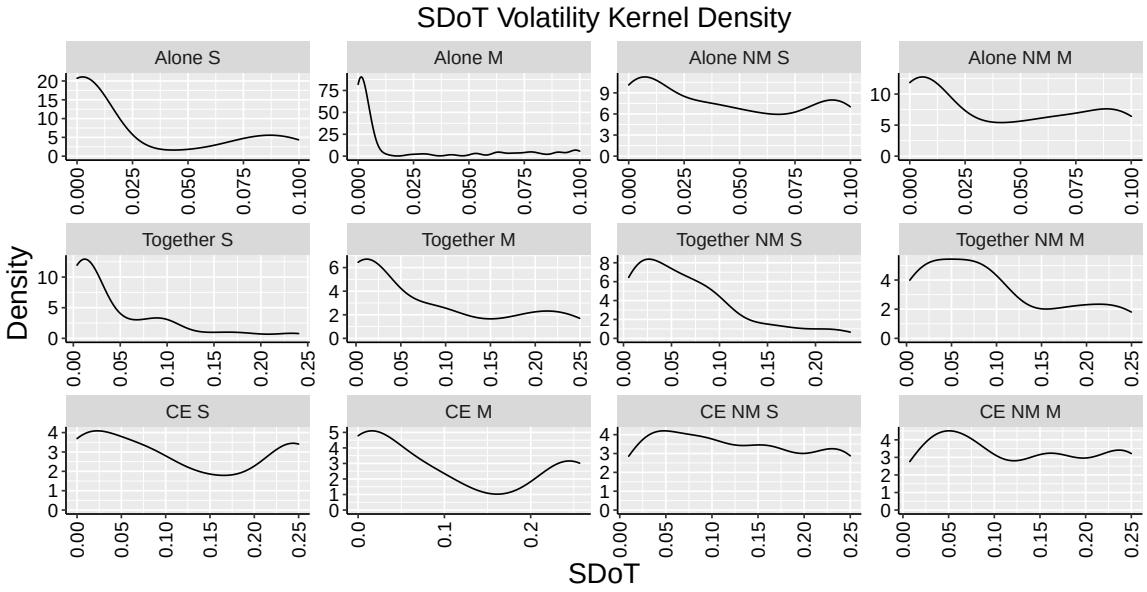


Figure 5.6: Kernel density estimations of the SDoT of agents that evolve alone, together, or with continued evolution (CE), and with or without neuromodulation (NM), to solve a single- (S) or multi- (M) stage task. *Note: graph scales are not comparable due to the disparity in the densities of each distribution.*

that evolution is more volatile for modulatory agents and that the received fitness tends to fluctuate often in agents that use neuromodulation. Further to this, agents have a lower mean and median CACoT and CCoT when evolving to solve a multi-stage task than a single-stage task, both with and without neuromodulation. The results therefore suggest that the best-in-population fitness fluctuates less often and by lower amounts during the course of evolution in agents that are tasked with solving a multi-stage task than those that solve a single-stage task. The exceptions to this are that: the mean CCoT of non-modulatory agents evolving together is higher for the multi-stage task than the single-stage task; the mean CACoT of non-modulatory agents evolving with continued evolution is higher for the multi-stage task than the single-stage task; and the CCoT of all agents evolving with continued evolution is higher for the multi-stage task than the single-stage task. A similar trend can be seen in Table 5.1, as more agents tend to solve the single-stage version of the task in each experiment than the multi-stage version; this would contribute to more fluctuations in fitness being observed during evolution, and more volatility as a result.

Non-modulatory agents have lower variability in CACoT and CCoT than their modulatory counterparts, however agents that use neuromodulation generally have a lower variability in SDoT than those that do not. These findings, combined with a lower mean and median in each metric, indicate that non-modulatory agents will be expected to have fewer

Metric	Exp	Task	NM	Moment				Median
				Mean	Variance	Skewness	Kurtosis	
SDoT	Alone	S	N	0.0251*	0.001 34	1.09	2.40	0.003 06*
			Y	0.0443	0.001 31*	0.254*	1.56*	0.0374
		M	N	0.0188*	0.001 01*	1.55	3.73	0.002 85*
			Y	0.0419	0.001 39	0.292*	1.49*	0.0382
	Together	S	N	0.0489*	0.003 71	1.65	4.94	0.0157 *
			Y	0.0665	0.003 15*	1.19 *	3.89*	0.0537
		M	N	0.0802*	0.007 35	0.763	2.07*	0.0310 *
			Y	0.0981	0.005 94*	0.636*	2.16	0.0804
	CE	S	N	0.118 *	0.009 65	0.236	1.47*	0.0956 *
			Y	0.127	0.006 29*	0.164*	1.70	0.111
		M	N	0.102 *	0.0104	0.507	1.56*	0.0616 *
			Y	0.129	0.006 90*	0.159*	1.59	0.133
CACoT	Alone	S	N	4.28 *	72.2 *	5.95	45.3	1.70 *
			Y	8.13	174	5.70 *	44.4 *	4.70
		M	N	2.79 *	18.0 *	3.25	14.7	1.10 *
			Y	4.37	29.5	2.57 *	9.59*	2.30
	Together	S	N	41.3 *	1730 *	1.55 *	5.18*	22.7 *
			Y	206	69 600	2.69	13.4	116
		M	N	40.8 *	6560 *	5.15	37.2	13.3 *
			Y	97.1	16 400	3.38 *	19.0 *	46.1
	CE	S	N	40.8 *	5480 *	3.04	13.9	40.8 *
			Y	317	250 000	2.68 *	11.1 *	317
		M	N	60.6 *	25 800 *	5.97	45.5	8.65 *
			Y	231	149 000	2.68 *	11.8 *	47
CCoT	Alone	S	N	19.9 *	1810 *	5.94	45.3	7 *
			Y	39.1	4340	5.70 *	44.4 *	22
		M	N	12.5 *	449 *	3.25	14.7	4 *
			Y	20.3	738	2.58 *	9.61*	10
	Together	S	N	155 *	39 600 *	2.52 *	11.6 *	75.5 *
			Y	854	1 610 000	2.85	14.1	356
		M	N	174 *	129 000 *	3.38 *	15.2 *	35 *
			Y	373	360 000	4.00	24.2	134
	CE	S	N	204 *	159 000 *	2.89	11.8	39.5 *
			Y	618	5 910 000	2.54 *	10.5 *	1620
		M	N	321 *	1 530 000 *	8.05	73.3	32.5 *
			Y	1130	3 470 000	2.58 *	10.9 *	228

Table 5.6: Statistical moments and median of the SDoT, CACoT and CCoT of the best-in-population agents after 500,000 generations of evolving alone or together, or after 1,000,000 generations with continued evolution (CE). Agents evolve to solve a single- (S) or multi- (M) stage task. The lowest values for agents with (Y) or without (N) neuromodulation (NM) are indicated with an asterisk (*).

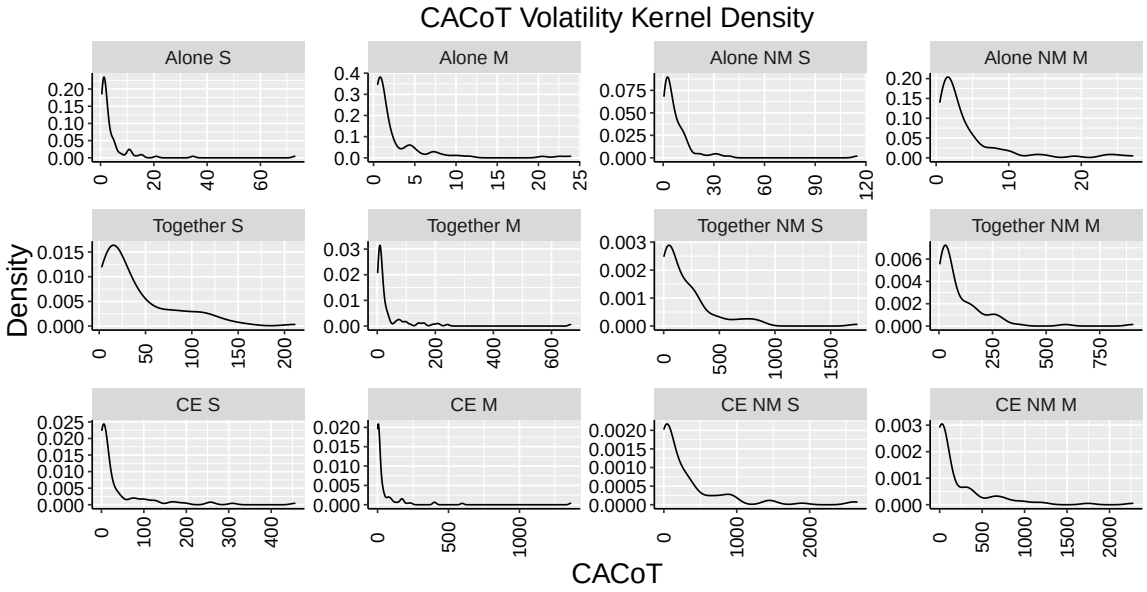


Figure 5.7: Kernel density estimations of the CACoT of agents that evolve alone, together, or with continued evolution (CE), and with or without neuromodulation (NM), to solve a single- (S) or multi- (M) stage task. *Note: graph scales are not comparable due to the disparity in the densities of each distribution.*

and more predictable fluctuations in fitness with less magnitude, and a higher and less predictable SDoT than those that use neuromodulation. Additionally, the mean, median and variance for each metric tend to increase as environmental variability increases. The results therefore suggest that agents experience more volatility as environmental variability increases, where volatility is likely to: be lowest in agents that evolve alone; increase when agents evolve together; and be highest when agents evolve with continued evolution.

Each metric for each experiment has positive skewness, showing that the data is right-skewed; this is supported by the median for each metric in each experiment generally being less than the mean. This skewness and the shapes of the distributions can be seen in the kernel density estimation plots for each metric, presented in Figures 5.6, 5.7 and 5.8. The distribution of values for the CACoT and CCoT of each experiment is highly skewed, whereas the SDoT is generally less skewed. Positive skewness indicates that agents would likely have an SDoT, CACoT or CCoT lower than the average, as the distribution is skewed by higher values; agents would therefore be expected to have a lower CACoT and CCoT than the observed mean and median. Further, the skewness and kurtosis of each metric is generally lower in modulatory agents than in non-modulatory agents; the values for each metric would be less likely to be extreme and more likely to be symmetrical around the mean, with outlier values being less likely in modulatory agents than non-modulatory agents.

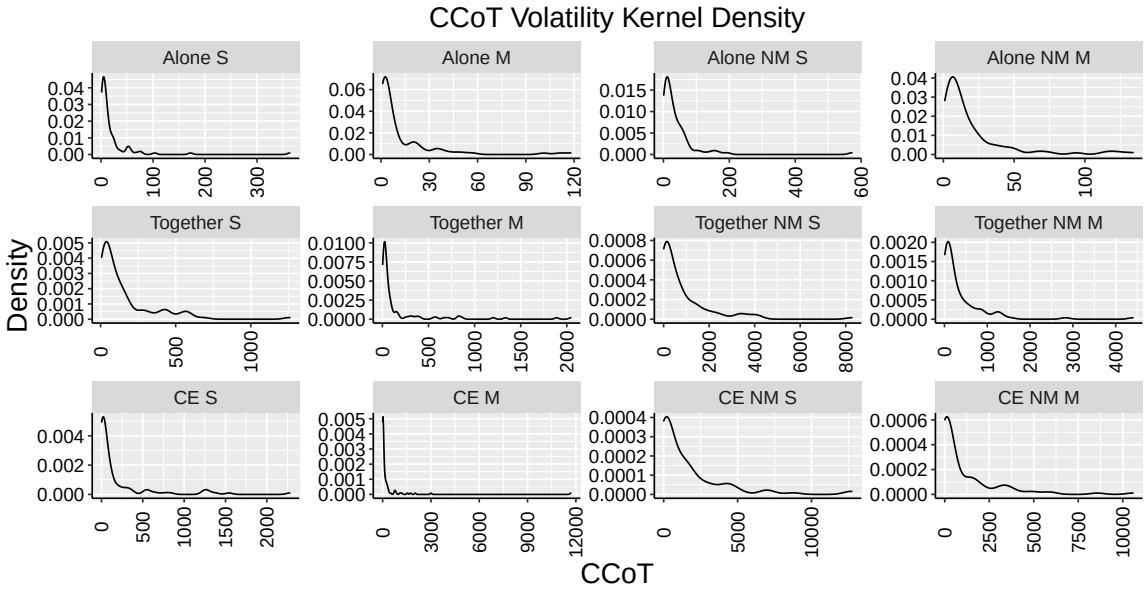


Figure 5.8: Kernel density estimations of the CCoT of agents that evolve alone, together, or with continued evolution (CE), and with or without neuromodulation (NM), to solve a single- (S) or multi- (M) stage task. *Note: graph scales are not comparable due to the disparity in the densities of each distribution.*

Statistical Tests – Comparing Non-Modulatory and Modulatory Agents

In addition to the analysis of fitness in Section 5.5.1, a Shapiro-Wilk test was conducted to detect normality in the SDoT, CACoT and CCoT distributions for each experiment; these results are presented in Table B.6 in Appendix B.2. The resulting p-value for each test was below 0.05, indicating that the distributions for each experiment are non-normal. As such, and in-line with the analysis performed in Section 5.5.1, Wilcox Signed Rank statistical tests (one two-tailed ($m_n \neq m_m$), and two one-tailed tests ($m_n < m_m$, $m_n > m_m$)) were then performed for each metric in each experiment; the results of these tests are presented in Table 5.7. The SDoT, CACoT and CCoT were each found to be significantly different between non-modulatory and modulatory agents ($p < 0.05$, $m_n \neq m_m$) in all experiments. There are however two exceptions to this. The first is that the SDoT of agents evolving with continued evolution to solve a single-stage task is not significantly different between modulatory and non-modulatory agents ($p = 0.376$); this might be due to similar numbers of agents successfully achieving their goal both with and without neuromodulation (Table 5.1), meaning that there is less difference between the recorded SDoT in each experiment (fewer agents do not achieve their goal than in other experiments, so there are fewer agents that receive a very small SDoT). Secondly, the SDoT of agents evolving together to solve a

Metric	Exp	Task (S/M)	Statistical Test Alternative Hypothesis			z	r
			$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$		
SDoT	Alone	S	1.297×10^{-4} *	6.484×10^{-5} *	9.999×10^{-1}	−3.829	−0.3829 (M)
	Together	S	4.535×10^{-3} *	2.267×10^{-3} *	9.978×10^{-1}	−2.840	−0.2840 (S)
	CE	S	3.760×10^{-1}	1.880×10^{-1}	8.129×10^{-1}	−0.8870	−0.088 70
	Alone	M	2.052×10^{-7} *	1.026×10^{-7} *	1	−5.196	−0.5196 (L)
	Together	M	6.919×10^{-2}	3.460×10^{-2} *	9.657×10^{-1}	−1.819	−0.1819 (S)
	CE	M	2.315×10^{-2} *	1.157×10^{-2} *	9.885×10^{-1}	−2.273	−0.2273 (S)
CACoT	Alone	S	5.881×10^{-5} *	2.941×10^{-5} *	1	−4.019	−0.4019 (M)
	Together	S	3.459×10^{-10}*	1.730×10^{-10}*	1	−6.278	−0.6278 (L)
	CE	S	5.787×10^{-9} *	2.893×10^{-9} *	1	−5.820	−0.5820 (L)
	Alone	M	3.982×10^{-4} *	1.991×10^{-4} *	9.998×10^{-1}	−3.543	−0.3543 (M)
	Together	M	1.371×10^{-6} *	6.856×10^{-7} *	1	−4.831	−0.4831 (M)
	CE	M	1.213×10^{-5} *	6.064×10^{-6} *	1	−4.377	−0.4377 (M)
CCoT	Alone	S	5.699×10^{-5} *	2.849×10^{-5} *	1	−4.027	−0.4027 (M)
	Together	S	3.520×10^{-8} *	1.760×10^{-8} *	1	−5.515	−0.5515 (L)
	CE	S	1.222×10^{-9} *	6.112×10^{-10}*	1	−6.080	−0.6080 (L)
	Alone	M	4.028×10^{-4} *	2.014×10^{-4} *	9.998×10^{-1}	−3.540	−0.3540 (M)
	Together	M	3.152×10^{-5} *	1.576×10^{-5} *	1	4.164	−0.4164 (M)
	CE	M	1.982×10^{-6} *	9.908×10^{-7} *	1	−4.757	−0.4757 (M)

Table 5.7: Wilcoxon Signed Rank statistical tests comparing the volatility metrics for the non-modulatory (m_n) and modulatory (m_m) agents in each experiment (Exp): evolving in an environment alone, together, or with continued evolution (CE), with a single- (S) or multi- (M) stage task. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

multi-stage task is also not statistically different between modulatory and non-modulatory agents ($p = 0.06919$); however, the SDoT was indeed found to be smaller in non-modulatory agents than modulatory agents ($p = 0.03460$, $m_n < m_m$). Overall, behavioural plasticity is found to only have a 'small', weak effect ($r = -0.1819$) on the SDoT of agents in the latter case. In all other tests, the SDoT, CACoT and CCoT for non-modulatory agents are found to be significantly lower than in modulatory agents ($p < 0.05$, $m_n < m_m$). This is supported by a negative effect size estimate r in each experiment.

Behavioural plasticity appears to have a stronger effect on the SDoT of agents that evolve alone compared to those that evolve together; the distributions of the non-modulatory and modulatory approaches diverge from each other more in single-agent environments than in multi-agent environments. This effect can be seen in Table 5.6, as the median SDoT for agents that evolve alone is more disparate between non-modulatory and modulatory agents, then when they evolve together. As the SDoT metric captures the standard deviation in fitness over time, this finding would indicate that more volatility is experienced during evolution when agents are situated in a shared environment than when alone; the actions

of others introduces more environmental uncertainty and unpredictability, which in turn decreases the strength of the effect of plasticity.

For the CACoT and CCoT, the effect that behavioural plasticity has on each of these metrics appears to be stronger in the multi-agent versions of each experiment compared to when agents evolve alone; additionally, the effect that plastic behaviour has on the volatility agents experience appears to be stronger when agents evolve to solve a single-stage task than a multi-stage task. This could be because fewer agents are generally able to achieve their goal when evolving to solve a multi-stage task than a single-stage task (Table 5.1).

Overall, this analysis shows that modulatory agents experience higher evolutionary volatility, where fitness fluctuates more often compared to their non-modulatory counterparts. This evolutionary volatility as captured by the SDoT, CACoT and CCoT metrics also tends to increase as environmental variability increases; in other words, as the environment gets more unpredictable and uncertain due to the unknowable actions of others, the received fitness over time for modulatory agents also tends to fluctuate more as a result. There does however seem to be a trade-off between fitness and volatility; despite this higher level of evolutionary volatility, modulatory agents are observed to have a higher mean fitness than non-modulatory agents (Tables 5.3 and 5.7), and achieve their goals more often.

5.5.4 Analysing the Modulatory Neurons in the Neural Networks

To understand the effect that behavioural plasticity can have on agents further, the arrangement of modulatory neurons that evolve in the agents across the study were examined to see whether any patterns emerge. For each of the 100 runs of each experiment, the deliberative network for the single best-in-population agent after evolution was recorded for comparison.

Table 5.8 presents the most common configuration of modulatory neurons evolved in the deliberative networks in each experiment, broken down into agents that do and do not achieve the goal. There is no single common configuration of neurons for the experiments where agents do not achieve their goal while evolving together or with continued evolution; this is because fewer agents do not achieve their goal than those that do. It is worth noting that the frequency of these common configurations is low compared to the total number of agents that do and do not achieve their goal (e.g. six agents had a common configuration out of 85 that achieved their goal when evolving alone to solve a single-stage task). As such, no configuration is common in all agents that achieve their goal across all experiments, and

Experiment	Task	Goal	L1	L2	L3	LT	Freq	Total
Alone	S	Yes	4	3	3	10	6	85
		No	3	3	3	9	2	15
	M	Yes	3	2	2	7	5	77
		No	3	2	3	8	2	23
Together	S	Yes	4	4	2	10	6	97
		No	-	-	-	-	-	3
	M	Yes	4	3	3	10	6	94
		No	-	-	-	-	-	6
Continued Evolution	S	Yes	4	2	1	7	7	100
		No	-	-	-	-	-	0
	M	Yes	5	4	2	11	8	99
		No	-	-	-	-	-	1

Table 5.8: The most common number of modulatory neurons evolved in each of the three layers of the deliberative neural networks (L1, L2, L3), and in total (LT). Results are presented for agents evolving to solve a single- (S) or multi- (M) stage task, and those that achieve their goal and those that do not. The frequency that the configuration occurs is shown, as well as the total number of agents overall. A dash (-) indicates that no configuration occurred more than once.

similarly no configuration commonly leads to agents not achieving their goal.

It is therefore apparent that agents can achieve their goal in many different ways, with different numbers of modulatory neurons in each layer and in different arrangements. It is not clear whether all modulatory neurons in these configurations are used or beneficial – some may be redundant if the surrounding weights are near zero values. Saying this, no agent was observed to evolve a neural network with either zero modulatory neurons or the maximum out of a possible 18 – each agent evolved a deliberative neural network with at least three modulatory neurons. This suggests that there is no obvious link between the number or configuration of modulatory neurons and: the success of an agent; the behaviours that the agent switches between; the stimuli that affects when modulation occurs; the type of environment it evolves in; or the task in which it has to solve. Because modulatory neurons can regulate neural network activity locally, this can potentially make goal-achieving behaviours (such as moving towards Water when a Stone is being carried) become accessible early on in evolution – without the agent needing to encode that exact knowledge directly in the network. This could be an explanation of why the mean best-in-population increases faster in agents that use neuromodulation than in non-modulatory agents in Figure 5.3. Further, agents did not converge to one single 'successful' or 'unsuccessful' configuration of modulatory neurons. Modulatory neurons can be arranged in a number of different ways to have a positive effect on agent evolution and fitness.

5.6 Discussion and Implications

Evolving to solve tasks in dynamic and uncertain environments can be a difficult process for neural networks, as learning or altering behaviours in response to changes in environmental stimuli means that knowledge encoded in the network will be changed. If this happens, behaviours that previously led to the achievement of goals may be lost, and fitness may therefore degrade as a consequence [78]. This becomes especially important to consider when designing real-world systems, as they can often be faced with solving tasks of different complexity, and with multiple stages [48]. In Chapter 4, the implications of evolving in a shared environment to achieve individual goals is demonstrated when evolving neural networks as agent controllers. When two agents – each unaware of the other – act within a shared environment, the consequences of these actions can interfere with the other's evolution and ability to achieve their own goals, making evolution more volatile. Unintended interactions can therefore have an unexpected impact on how well suited an agent or system is for the environment it is located in. These issues are becoming evermore important to consider when designing technical systems, as they are increasingly being composed of many components or sub-systems; as a result, it becomes increasingly likely that these systems will interact in unintended and unpredictable ways [101], which can lead to the state of the environment changing without warning through the actions of others. The ability for a system to behave appropriately regardless of unexpected interactions or changes in environmental states therefore becomes crucial.

In nature, behavioural plasticity is one mechanism that helps animals and humans to temporarily change their behaviour in an attempt to survive or overcome environmental changes. In this study, this concept is abstracted by exploring how simulated agents evolve in both single- and multi-agent environments, when agents evolve to solve either a single- or multi-stage task. Here, the variability of each experiment differs in terms of whether there is another agent that can affect the state of the environment with its actions. In all cases, agents have no knowledge of other agents in their environment and therefore cannot intend to interact with others – nor can they evolve to understand that a perceived environmental change is caused by the actions of another agent.

Behavioural plasticity in the form of evolving with neuromodulation is shown to have a positive effect on the fitness that agents receive throughout and at the end of evolution.

However, what may not be so intuitive is the observation that behavioural plasticity also increases the volatility of agent evolution: whilst agent fitness is higher in modulatory agents than non-modulatory agents overall, the fitness over evolution fluctuates more. One might expect that dynamic behaviour would – in addition to improving the chance of success – actually decrease the amount of volatility within the system, by counteracting any dynamics or volatility present within the environment. The three metrics used to measure the amount of evolutionary volatility (SDoT, CACoT, and CCoT, Table 5.6) show this is not the case, as the mean and median for each metric tends to increase as the variability in the environment increases. Even when an agent evolves alone – and therefore doesn't experience interference – agents that are capable of behavioural plasticity experience more evolutionary volatility than agents that are not.

Agents that evolve alone to solve either a single- or multi-stage task experience less evolutionary volatility than agents that share an environment with another agent. These results corroborate the findings presented in Chapter 4, which demonstrates that the actions of another, unknown agent within the environment can make the environment appear unpredictable and variable; this in turn increases the evolutionary volatility of agents in shared environments, since agents may attempt to change their behaviour when experiencing changing environmental stimuli. Whilst the overall number of successful agents increases when agents are in shared environments compared to when they are alone, the number of agents able to achieve their goal individually in fact decreases; this shows that it is harder to evolve individual behaviour when the environment is prone to change due to the actions of others. More agents overall are able to achieve their goal in a shared environment than an individual environment, as exploitative behaviour can emerge; these exploitative agents evolve to rely on the actions of their unknown partner agent in order to achieve their goal. From an agent's perspective, the environment becomes predictable because agents are unable to perceive others or their actions, and thus there is no incentive to evolve individual goal-achieving behaviours as this comes at a personal cost. To demonstrate this further, when an agent first evolves alone and then continues to evolve in a shared environment, the change in context and increase in variability also affects the individual goal-achieving behaviours evolved when alone; when modulatory agents continue to evolve with another agent, some agents lose their individual goal-achieving behaviour in favour of exploiting the other agent to receive a higher fitness. Whilst this is beneficial for the agent in terms of the

fitness received, agents are reliant on events beyond their control in order to achieve their goal. If an agent was to suddenly find itself in an environment alone again, it is likely it would have to rediscover the goal-achieving behaviours it once possessed.

The findings and analyses presented in this chapter show that consequences can exist for systems that are unable to act appropriately in environments that are prone to change, or when there is the potential to be affected by the actions of an unknown other system in the same environment. In reality, as systems and the components they are comprised grow larger, the opportunities for unintended interactions with others and unexpected environmental changes also increase. Behavioural plasticity is one route to equipping systems with the ability to overcome environmental uncertainty, although this is shown to lead to a higher level of evolutionary volatility. However, higher volatility in this case is the cost of an increase in fitness and ability to achieve goals; agents spend more of their lifetime or evolution able to achieve their goals than those that are not capable of behavioural plasticity, which provides further evidence of the benefit of plasticity. A trade-off between fitness and volatility therefore exists in plastic agents, where the benefits of plasticity are demonstrated in both single- and multi-agent environments, where agents evolve to solve different complexities of task. This study demonstrates that designers of systems have a choice between predictability in terms of the fitness achieved over time, or an increase in fitness that may come as a result of exploitative behaviour. Behavioural plasticity may therefore not be entirely suitable in environments where an agent relying on the actions of another to achieve goals is not desirable; the following chapter will examine the effects of plasticity and increasing variability on agent evolution in closer detail.

5.7 Conclusion

In this chapter, the River Crossing Dilemma testbed is used to explore the effect that activity-gating neuromodulation has on an agent's ability to evolve in environments when experiencing interference. Additionally, the Protected River Crossing Dilemma testbed is introduced, which incorporates a less complex, single-stage task compared to the multi-stage task seen in the RCD; this is so that agent evolution can be analysed when agents evolve to solve different complexities of task. An important element of this study is that agents do not have the capacity to know of or learn about the actions or existence of other

agents; in this way, they cannot intend to cooperate or exploit one another.

The main findings and contributions of this chapter are:

- Behavioural plasticity benefits agents that evolve to solve different complexities of task in environments that vary to different extents, as plastic agents achieve their goals more often and achieve higher fitnesses as a result. This relates to the third research question outlined in Chapter 1.
- Variability arising from the actions of other, unknown agents in an environment can be overcome by using behavioural plasticity. This relates to the first and third questions outlined in Chapter 1.
- Behavioural plasticity increases fitness as well as the evolutionary volatility experienced by agents in tasks of varying complexity. This relates to the second and third questions outlined in Chapter 1.
- Evolving individual goal-achieving behaviour is harder in multi-agent environments, however agents may evolve to exploit the actions of others by sacrificing their ability to achieve their goal individually. This relates to the first question outlined in Chapter 1.
- Modulatory agents are shown to achieve their goals in many different ways – there is no single arrangement of modulatory neurons that influences an agent's ability to achieve goals in any of the experiments conducted. This relates to the third research question outlined in Chapter 1.

These results demonstrate that behavioural plasticity as a result of activity-gating neuromodulation has an observable effect on the expected fitness of evolved agents in different environments; this behavioural plasticity is beneficial to create adaptive agent controllers that can temporarily and reversibly change behaviour in novel situations. Specifically, plasticity is beneficial to agents that exist in individual or shared environments, as well as those that solve simpler single-stage, or more complex multi-stage tasks. Additionally, when the context of an agent's environment changes from being individual to shared, neuromodulation helps agents to adapt and succeed to the new context and change in environmental stimuli.

Using three metrics to analyse evolutionary volatility, agents exhibiting behavioural plasticity as a result of neuromodulation are shown to experience more fluctuations in fitness

during evolution than those that do not showcase plastic behaviour; this higher volatility is a result of modulatory neurons regulating activity in the neural networks, as a response to changing environmental stimuli. While experiencing a higher level of evolutionary volatility is a consequence of the ability to temporarily and reversibly demonstrate different behaviour by using neuromodulation, agents are more likely to achieve their goals, spend more of their lifetime or evolution being able to achieve their goals, and receive a higher fitness than non-modulatory agents. Behavioural plasticity arising from neuromodulation therefore creates a trade-off, as a significantly higher expected fitness and chance of goal-achievement comes at the cost of higher evolutionary volatility. This may indeed be desirable for agents that exist in highly unpredictable and unknown environments, by equipping them with the ability to respond quickly and appropriately to environmental change in a way that preserves or even improves fitness or performance.

This study has explored the impact of behavioural plasticity on evolutionary volatility and agent fitness, by comparing how agents evolve in single- and multi-agent environments to solve single- and multi-stage tasks. The most variable environment in this study evolved agents alone for an initial period of time, before continuing to evolve with another agent; the study presented in Chapter 6 will investigate the extent to which behavioural plasticity enables agents to maintain their goal-achieving behaviours when the presence of other agents in the environment is unpredictable and uncertain. This would create more uncertain conditions for agents to evolve in than those observed in this chapter.

As systems are increasingly located in unpredictable and variable environments, possessing the ability to overcome the unknown and to behave appropriately in unseen scenarios is evermore important. This study demonstrates that activity-gating neuromodulation allows agents to temporarily and reversibly change behaviour in response to changing environmental stimuli without affecting knowledge encoded in their neural networks. Whilst behavioural plasticity is shown to improve fitness during agent evolution and the ability to achieve goals, a trade-off is also shown to exist as agents experience more volatility during evolution as a result.

Chapter 6

Environmental Variability and Evolutionary Volatility

The work presented in this chapter has been adapted from the following publication:
[20] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Evolving Neuromodulated Controllers in Variable Environments. In *Proceedings of the IEEE 2nd International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, 2021. In Press.

6.1 Motivation

In the natural world, both humans and animals alike are often situated in environments that are dynamic, uncertain, and prone to changes in context over time. One way that organic creatures have adapted to survive in their variable environments is through exhibiting plasticity; this ability to temporarily and immediately alter phenotypic traits like behaviour depending on the environmental stimuli helps to overcome the unknown [196, 122, 59]. As discussed in Chapter 5, behavioural plasticity can be achieved through neuromodulation, a biological process found within the brain whereby synaptic activity is regulated in response to novel or changing stimuli [1], resulting in short- or long-term changes in behaviour. The decision-making processes that we, as humans, and animals make must therefore be adaptable over time in order to behave appropriately in response to environmental changes, to achieve goals and ultimately to survive [120, 177].

The experimental study presented in this chapter extends that of Chapter 5, with the intention of exploring how behavioural plasticity may affect goal-achievement when artifi-

cial agents repeatedly find themselves in new situations. This is henceforth termed 'environmental variability', referring specifically in this study to the changes in environmental conditions that evolutionary agents face either during or between generations. As established in Chapter 4, the actions of other agents in the environment can cause interference, leading to changes in the way that agents evolve; if the effects of such interference are not understood, agents will perceive the environment to be changing of its own accord, thus creating new and unpredictable situations for them to deal with. Further, how agents may evolve to express 'intelligent behaviour' [231] when encountering these unforeseen conditions in environments that are highly variable is explored, as they experience a range of scenarios and utilising this knowledge to succeed in the face of the unknown. By evolving agents in increasingly variable environments, agents can potentially experience a diverse range of environmental conditions; one would imagine that the more knowledge that an agent gathers about past experiences, the more resilient it would be to new contexts, situations, or changes in environmental stimuli. One of the aims of this chapter is to therefore understand how agents may mitigate interference from other agents in highly variable environments.

Komers [122] states that the degree of plasticity would increase with variability in the natural world. The focus of this chapter is to therefore explore whether a relationship between variability and plasticity is also observed in artificial agents. Specifically, this focus is on observing how agents with no capacity to perceive others evolve in environments with different levels of variability, to analyse the effect that neuromodulation has on agents in these increasingly uncertain environments. In Chapter 5, behavioural plasticity is shown to be a useful mechanism for survival when agents are faced with this kind of environmental uncertainty – at the cost of evolutionary volatility; as such, the effect that neuromodulation has on agent evolution and goal-achievement is also explored, to ascertain the extent to which temporary, reversible changes in behaviour enables agents to achieve their goals in increasingly variable environments. Further, the evolutionary volatility of these agents is measured to ascertain whether a trade-off between fitness and volatility exists when agents evolve to achieve individual goals in dynamic environments.

The remainder of this chapter is organised as follows: Section 6.2 discusses how environmental variability has been explored within the literature; Section 6.3 details how the experimental study that this chapter is based upon is designed, using the River Crossing Dilemma testbed introduced in Chapter 3; Section 6.4 then presents the results of the study,

which is analysed further in Section 6.5; finally, Section 6.6 discusses the findings and their implications, with Section 6.7 closing the chapter by stating the contributions made.

6.2 Exploring Plasticity and Variability

The study in this chapter extends that which is presented in Chapter 5, where the effect of behavioural plasticity is explored in agents that evolve in increasingly variable environments. To elaborate on discussion of related literature in Section 5.2, this section firstly reiterates the general notion of behavioural plasticity; the literature surrounding environmental variability and dynamicity is then explored, to provide a solid grounding for the environmental study detailed in Section 6.3.

6.2.1 Behavioural Plasticity via Neuromodulation

Behavioural plasticity is observed in the natural world in animals such as the orb spider, *Eustala illicita* [106], the Namibian rock agama, *Agama planiceps* [49], and the African striped mouse *Rhabdomys pumilio* [185]; these organisms can change their behaviour in response to changing environmental stimuli, and novel or variable environments [199, 196]. Behavioural plasticity is thought to increase the chances of survival of organisms that exist in novel, dynamic or uncertain environments [196], enabling rapid reactions to unknown stimuli. If an organism experiences a change in environmental stimuli, or if the environment is in a constant state of flux, 'activational plasticity' [196] can allow the organism to alter its behavioural responses in a short-term, immediate, and reversible way [122, 90, 69]. These behavioural changes can be achieved with *neuromodulation*, which strengthens or weakens the synapses between neurons; this results in the regulation of synaptic plasticity based on the chemical modulatory signals, the pre- and post-synaptic activities, and the environmental stimuli [1, 2]. Neuromodulation can either result in learning in the long-term if synaptic strengths are changed over time, or immediate, temporary behavioural changes if a modulatory signal temporarily changes the synaptic activity – the latter of which is the focus of this chapter and that which precedes it.

6.2.2 Variable Environments and Plasticity in the Natural World

Komers [122] states that although learning is one type of behavioural plasticity, it is not a prerequisite for plastic behaviour; instead, 'innate' changes to behaviour, described above as 'activational plasticity' [196] are sufficient. Further, Komers expects that the degree of behavioural plasticity would increase with environmental variability, implying that organisms capable of temporary behavioural changes are better suited to, and change their behaviour more often when situated in variable environments. In one study that looks at how environmental variability affects the behaviour and rehabilitation of hatchery reared cod, Braithwaite and Salvanes [42] found that the development of 'flexible' or plastic behaviour is promoted when the fish are placed in environments prone to change; this was found to increase the chances of survival and successful rehabilitation in the wild, whereas fish in the least variable conditions performed poorly from a lack of adaptability. Even the nematode *C. elegans* is found to exhibit plasticity by altering future behaviour if environmental stimuli is perceived to have changed. *C. elegans* contains only 302 neurons, which pales in comparison to the 86 billion neurons that the brain of an adult human male contains on average [11]; the nematode is thus a well-suited candidate for studying neural mechanisms, such as those that underpin plasticity, in close detail [42]. These findings support the expectations of Komers [122], in that plastic behaviour is both prevalent in organisms in highly variable environments, and less prevalent in more predictable environments.

Further to the studies of the effect of variability in the natural world described above, human behaviour has also been studied in the context of usage of common-pool resources, however this is moreso in terms of social dynamics rather than plasticity. Anderies et al. [7] show that in an 'irrigation game', which involves physical infrastructure to manage a common-pool resource and managing an asymmetrical common-pool resource dilemma for access to the resource (where water is the resource in this case), participant cooperation drops in variable environments compared to a low-variability environment. Additionally, when an environment shifts upwards in the level of variability, understanding of the structure of the system is said to play a part in the performance of the participants and their ability to adapt to variable environments [7]; this can be thought of in the context of plasticity, in that a better understanding of variability in the environment and its consequences can lead to more plastic and successful behaviour. On the other hand, Fehl et al. [83] found

that cooperative behaviour is more prevalent in dynamic networks (i.e. those where partners can change) of anonymous human participants playing iterated Prisoner's Dilemma games, compared to static networks of fixed partners. An interesting question to explore in the experimental study presented in this chapter would therefore be whether the level of environmental variability would have an impact on the social dynamics that emerge in the environments – especially as agents have no knowledge of others, or their intentions, actions, or capabilities.

In this chapter, as well as the preceding study presented in Chapter 5, the concept of regulating *behaviour* rather than learning is explored; as such, this chapter explores the effect of activational plasticity – the immediate and reversible changes to behaviour, facilitated by neuromodulation – when agents evolve in environments that vary to different extents, both within and between generations. It would therefore be expected that any observed benefit of behavioural plasticity is sustained as environmental variability increases, as plastic organisms in nature are observed to have an increased degree of plasticity in more variable environments [11, 42, 122].

6.2.3 Variability in Artificial Environments

In the context of agent-based systems, when an agent encounters a new context, situation, or environment, past experience or knowledge may not be either useful or beneficial when the environment has elements of dynamicity or uncertainty [216]. Cheng et al. [54] discuss the challenges faced by technical systems operating in real-world environments shared with other systems (in this instance, robots); namely, these include interacting or behaving appropriately despite the heterogeneity of other systems or obstacles within the environment, as well as the assumed perfect knowledge of the capabilities or actions of others. In reality, such wealth of knowledge is an expensive and often unrealistic luxury. In nature, if full information about the environment, the consequences of the actions of others, or even the presence of others is not known, evolution can be negatively affected [122]. This however is becoming increasingly commonplace in heterogeneous technical systems, as components within the system may not initially have the ability to perceive, learn about or interact with others [5].

The way in which agents learn or evolve in variable environments has been explored in the domains of reinforcement learning [184, 158], social learning [158, 116] and robotics [50,

54], to name a few. Carvalho and Nolfi [50] define the 'behaviour' of agents in variable environments as the process arising from agent or environmental interactions; specifically, the authors show that robots with neural controllers can evolve behavioural plasticity over time, enabling them to achieve their goal in a room-cleaning task more efficiently when presented with changing environments. Carvalho and Nolfi [50] also express the importance of behavioural plasticity as a mechanism that can be used to overcome variability in both the internal and external environment, stating that:

> "*...all environments are variable, from the perspective of an organism that is situated and performs actions in an environment, independently of whether they appear variable or not from the perspective of an external observer.*" [50]

Ndousse et al. [158] also explore how agents evolve in variable environments, however the effect of social learning in this context is explored instead of plasticity; variability in that study arises from being situated in either a single- or multi-agent environment, and the changes in effects that the presence of a partner can have within the environment itself. The authors found that the opportunity to gain experience through social learning actually improves agent performance in solitary environments, compared to agents that only have the chance to learn alone. Learning from a broad range of experiences and contexts in a multi-agent environment appears to be beneficial when the context of the environment changes such that an agent finds itself alone.

Drawing the focus back to how this discussion of variability relates the study in this chapter, the original River Crossing Task [178], which inspired the design of the RCD testbed (Chapter 3), could also be said to incorporate elements of variability within the environment. Not only does the environmental configuration change for agents at every generation, but agents are also evaluated on three consecutive environments of increasing difficulty, in terms of the width of the river and consequently the number of Stones required to build a bridge. Developing on from this, Borg et al. [37] increase the difficulty of the environment further in the RC+ task, where the availability of Stones within the environment decreases as river width increases, such that the fifth and final environment cannot be solved by building a bridge. As a result, in both of these studies, agents must evolve behaviours that are robust to the variability encountered within the environment in terms of the physical attributes of the environment, and the behaviours required in order

to complete the task itself. Whilst variability has been incorporated into other instances of the River Crossing family of testbeds, such as the aforementioned RCT and RC+, the consequences of variability in multi-agent environments has not yet been explored, nor in the context of behavioural plasticity.

The intention of this chapter is to explore how evolutionary agents without the ability to learn of or perceive others within the environment are able to evolve to achieve their goals in varying environments. Variability can arise from the state of the environment itself, the actions of others within the environment, or even the existence of others; therefore, if the presence of another agent or the consistency of its actions changes, variability would increase. The effect of behavioural plasticity in the presence of variability is thus observed, to ascertain how it may help these agents to achieve their goals in environments in which another agent may or may not exist. Specifically, the study is intended to investigate how behavioural plasticity via neuromodulation can affect agent evolution in increasingly variable environments. In terms of the River Crossing Dilemma testbed in which the study in this chapter is executed, agents will perceive the environment to increase in variability if the actions of another agent change the state of the environment; as such, the availability of Stones and the subsequent construction of bridges – where the former are necessary objects for the latter – can vary from environment to environment due to interference.

6.3 Experimental Study

The experimental study presented in this chapter is designed to explore how activity-gating neuromodulation affects the evolution of agents that are situated in variable environments, and how the level of variability that an agent experiences can in turn affect evolution and goal-achievement. In Chapter 5, behavioural plasticity is shown to increase the likelihood that agents will achieve their goal when evolving to solve a single- or multi-stage task, but this benefit comes at the cost of evolutionary volatility; here, the evolutionary volatility of agents evolving in variable environments is analysed to ascertain whether a similar trade-off exists when the environment is prone to change.

6.3.1 Testbed Design

This experimental study uses the RCD testbed, described in Section 3.2 and Figure 3.1, to explore how agents evolve to achieve their own goals when their environment is unpredictable and prone to change. One of the limitations of the study presented in Chapter 5 is that agents evolve in a single environment, and exist either alone, with another, or first alone and then with another; this means that an agent could potentially evolve to depend on the actions of another agent in order to achieve its goal, since the environment becomes almost dependable over time. What isn't however addressed is how agents evolve when agents are faced with having to achieve goals in contexts where another agent may or may not exist. The experiments detailed in this section are designed to study how environmental variability affects agent evolution, where variability is affected by: (1) the number of environments that agents are evaluated on at each generation, (2) whether an agent exists alone or if another exists within the environment, and (3) the dynamicity of other agents, which impacts the reliability of one's actions within the environment. The degree of plastic behaviour is said to increase in line with environmental variability [122]; as such, it would be expected that plastic agents would be capable of expressing a range of behaviours in response to changing environmental stimuli, thus increasing chances of survival and goal-achievement in comparison to non-plastic agents. The specific details of the experimental design are discussed in Section 6.3.3.

Number of environments: The first way in which the level of environmental variability differs between experiments arises from the number of environments that agents are evaluated on at each generation. In previous chapters, agents are evaluated on one environment at each generation to observe how agents evolve either alone or in a shared environment; other studies that use instances of the River Crossing family of testbeds for example evaluate agents on a series of environments that increase in difficulty. Robinson et al. [178] evaluate agents on three consecutive environments of increasing difficulty in the original River Crossing task; Borg et al. [37] however increase the difficulty further by evaluating agents instead on five environments in the RC+ task, such that the final environment cannot be solved by building a bridge. A question arising from the design of these studies is thus *how does the number of environments that agents are evaluated on affect agent evolution?* This study is designed to explore this question by evaluating agents on a different number of en-

vironments in each experiment; this alters the difficulty of the experiments, but further, the differences between each consecutive environment that agents are evaluated on contribute to the notion of 'environmental variability' discussed in this chapter. In other words, agents would face more variability as the number of environments they are evaluated on increases.

Presence of another agent: In previous chapters, agents are situated in an environment in which they exist alone, share with an unknown other agent, or evolve alone for an initial period of time before continuing to evolve in a shared environment (termed 'continued evolution'). Thus far, the evolution of agents in shared environments with no ability to perceive or reason about other agents is shown to be affected by the indirect actions of other agents *through* the environment. A question arising from these experiments is *are agents able to retain goal-achieving behaviour if the presence of another agent varies?* Ndousse et al. [158] investigated a similar concept, finding that social learning is beneficial when reinforcement learning agents are trained in both single- and multi-agent environments; social learning is specifically shown to improve an agent's performance when existing alone in an environment, compared to those that are only trained alone. Behavioural plasticity is the focus of this study instead of social learning, however it would be interesting to see if similar results would be observed when agents are evolved in the RCD in both single- and multi-agent environments; further, agents in this study have no ability to learn *of* others, so cannot consequently learn *from* others. The second way in which the level of environmental variability changes between experiments thus emerges from evaluating agents in environments where another agent may or may not exist. In varying the presence of another agent, indirectly depending on the actions of others (in terms of exploitative behaviour) is a risk because the agent may not always exist in a shared environment; because agents cannot perceive other agents within the environment or reason about their actions, the unpredictable presence of another agent would appear as variability arising from the environment itself.

Dynamicity of partners: Sharing an environment with another agent unlocks the potential for emergent cooperation or exploitation; in this study, the behaviour leading to these social dynamics however cannot be intentional as the agents have no ability to perceive or learn of other agents within the environment. Further, the actions of another agent can alter the environment in an unknowable way, which can change the reliability of an agent's own

actions within the environment – the effect of which has been demonstrated in earlier chapters. A final question arising here is *how does the dynamicity of partner agents affect agent evolution, and the emergent social dynamics that may arise?* If the actions of other agents in a shared environment are prone to change, environmental variability would increase; agents would be incapable of understanding or perceiving that environmental changes are the direct result of the actions of other agents – they would instead perceive the environment *itself* to be changing unpredictably. In the domain of human behaviour and cooperation, Fehl et al. [83] found that cooperative behaviour is more prevalent in dynamic networks (i.e. those where partner can change) of anonymous human participants playing iterated Prisoner's Dilemma games, compared to static networks of fixed partners. The third method of altering variability in the experiments therefore arises from the perceived dynamicity of the actions of another agent; when an agent shares an environment with a partner agent, that partner will either *coevolve*[1], or will be *random* at each generation. Sharing an environment with random partners throughout evolution would increase the perceived variability of the environment compared to a coevolved partner, because previously successful behaviour of an agent (where 'success' is in terms of achieving the goal) may become unsuccessful or unreliable; each random partner has the potential to affect the environment in a previously unseen way. It would be interesting to see whether unintentional cooperative or exploitative behaviour is favoured during evolution in these experiments in line with the findings of Fehl et al. [83], or whether the ability to perceive or learn that partners may be coevolved or random would be needed for these behaviours to emerge.

6.3.2 Agent Design

Aligning with the experimental study designed to explore behavioural plasticity via neuromodulation in Chapter 5, the agents in this study employ the same two-tiered neural network architecture that was introduced in Section 5.3.2, which can also be seen in Figure 5.1. Further, agents capable of behavioural plasticity operationalise neuromodulation as described in Sections 5.3.3 and 5.3.4, to remain consistent with the previous study.

[1]Note that 'coevolve' here is used to describe two agents with distinct genetic material from two separate populations, that evolve in a shared environment. This is different to the 'coevolutionary optimisation' as discussed in Section 2.6, which refers to individuals whose fitnesses are closely coupled and dependent on one another; whilst the fitnesses of agents in these experiments can be *influenced* by the actions of others as a result of interference, both agents are independent and can achieve their goals in isolation or in shared environments.

6.3.3 Experimental Design

Each experiment is repeated 100 times using 100 seeds; agents are evolved either with or without neuromodulation for 500,000 generations in the RCD testbed, and are randomly initialised. As with the studies in previous chapters, agents in these experiments do not 'learn' within their lifetime; if an agent is evaluated on multiple environments at each generation, its genetic code (i.e. the weights and configuration of neurons in the deliberative neural network) remains unchanged. 'Learning' therefore only occurs in the breeding process between generations, through mutations – not during the evaluation process at each generation. These experiments are summarised in Table 6.1.

Evolving with another agent: The first set of experiments explores how agents evolve in a shared environment; agents are evaluated on one environment at each generation. The environment is more variable and unpredictable when agents evolve with random partners than a coevolved partner; each random partner will act differently within the environment, therefore allowing the agent to learn from and generalise to more scenarios compared to evolving with a single, coevolved partner.

Evolving when alone or with another: An issue with only evolving in a shared environment is that agents may evolve to depend on other agents in order to achieve their goal. Specifically in the case of the RCD, if an agent evolves to exploit another agent and doesn't learn to build a bridge (the necessary behaviour to cross the river and collect its allocated Resource, therefore achieving the goal individually), then it may be unable to achieve its goal if it then finds itself alone. The second set of experiments is designed to bridge this gap by evaluating agents on two consecutive environments at every generation. Firstly, agents are evaluated in an environment where they exist alone; once the first evaluation is terminated (either by completing or failing the task, or reaching the maximum amount of moves without success or failure), the environment is reset. Agents are then evaluated in a shared environment, with either a coevolved partner or random partners. By evolving agents in both a single-agent and a multi-agent environment at each generation, they have the opportunity to evolve appropriate behaviours that are successful whenever they are alone or with a partner.

Section	Partner Type	Number of Environments	Description
6.4.1 6.4.2	Coevolved Random	1	Shared
6.4.3 6.4.4	Coevolved Random	2	Alone, Shared
6.4.5 6.4.6	Coevolved Random	4	Alone, Shared, Alone, Shared

Table 6.1: Breakdown of the experimental study presented in Chapter 6.

Evolving in highly variable environments: The final set of experiments extends this notion by evaluating agents on four consecutive environments at each generation. Agents exist alone in the first and third environments, whereas the second and fourth environments are shared with coevolved or random partners. Regardless of the dynamicity of the partner agent, the partners in the second and fourth environments are different to each other; agents will either coevolve with two separate partners in the second and fourth environments, or two different random partners, depending on the dynamicity of the experiment. It may seem as though evaluating an agent for the second time where it is alone is redundant, however as agents are able to act differently in each environment (i.e. they do not follow a predetermined plan of actions or path, because their next move is dictated by their current state at each timestep and an element of stochasticity), designing the experiments this way means that agents can learn from more experiences. Agents are thus able to draw on more experiences through being evaluated on four environments at each generation, as opposed to three (e.g. one single-agent, and two multi-agent environments).

The overall generational fitness for an agent is the sum of the fitnesses achieved in each individual environment it is evaluated on, as calculated with Equation 3.1; this means that the maximum fitness for agents evolving in one, two and four environments is 1.0, 2.0 and 4.0 respectively. In reality, the maximum fitness an agent can achieve when alone is 0.7 as it must endure the total cost of building a bridge in order to achieve its goal, so the actual maximum fitnesses for each experiment in this study are 1.0, 1.7 and 3.4 respectively. It is also worth noting that when an agent has a random partner at each generation, for all 100 repetitions of the experiment agents are paired with the *same* generational partners to ensure consistency and comparability in the results. However, when an agent coevolves with another, these partners differ for each of the 100 repetitions so the results are not biased towards one specific partner agent.

Partner	NM	Fitness (% of Runs)				Goals Achieved		
		0.7 (Indep)	0.9 (Coop)	1.0 (Exploit)	≥ 0.7 (GAF)	0	1	$\bar{x}$
Coevolved	No	27	5	36	68	32	68	0.68
	Yes	44	0	50	94	6	94	0.94
Random	No	2	29	29	60	40	60	0.60
	Yes	3	59	24	86	14	86	0.86

Table 6.2: The percentage of modulatory and non-modulatory agents across 100 runs that achieved their goal with common fitnesses after evolution. Agents are evaluated on one environment with coevolved or random partners. A fitness of 0.7 is achieving the goal independently by building a bridge with two Stones (Indep); 0.9 is cooperation (Coop) by sharing the cost of bridge-building; 1.0 is exploitation (Exploit); ≥ 0.7 is a goal-achieving fitness (GAF).

6.4 Results

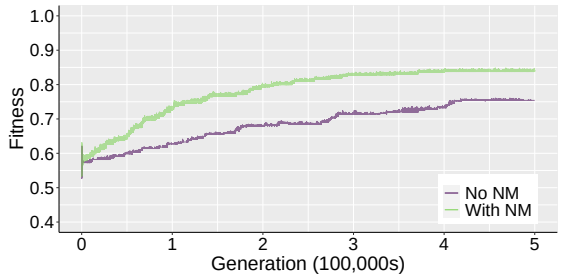
6.4.1 Evolving with a Coevolved Partner

To begin the study, the impact that sharing an environment with another agent can have on the way an agent evolves to achieve its own individual goals is explored; here, agents are coevolved with a single partner agent. As these both of these agents evolve in tandem, they would face the least amount of uncertainty and environmental change during evolution and would consequently gain little experience or exposure to different environmental conditions. These results are directly comparable to those in Section 5.4.5, as agents evolve in an environment with one partner to learn a multi-stage task in the RCD.

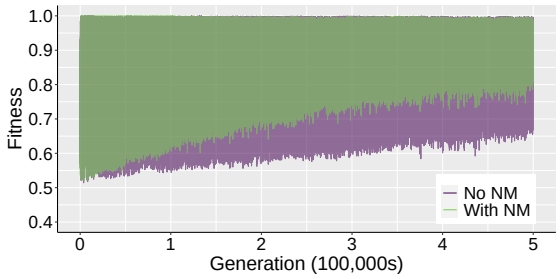
Figure 6.1a shows the mean best-in-population fitness over time for agents evolving both with and without neuromodulation. Agents that display behavioural plasticity by evolving with neuromodulation are observed to have a higher mean best-in-population fitness than those that do not – an effect that is sustained throughout evolution. Further, 68% of non-modulatory agents receive a goal-achieving fitness at the end of evolution compared to 94% of modulatory agents, when they are evaluated on one environment at each generation with a coevolved partner (Table 6.2). This indicates that more agents are able to achieve their goal with neuromodulation than without.

6.4.2 Evolving with Random Partners

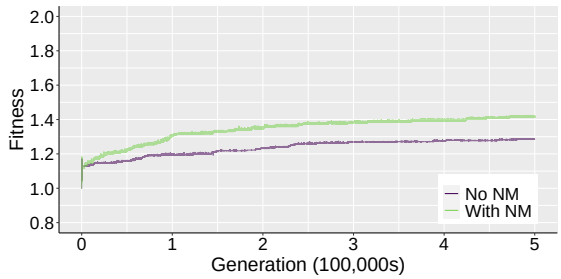
These experiments explore how evolving in an environment where the other agent changes at each generation can affect agent evolution and the ability to achieve goals. Unlike the experiments discussed in Section 6.4.1, each random partner has the potential to act within,



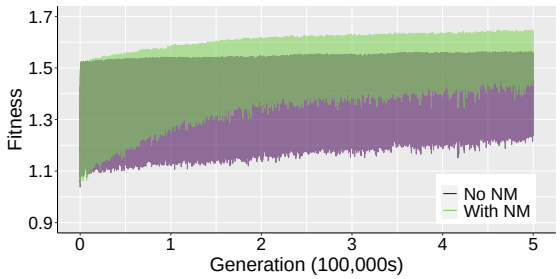
(a) One Environment with a Coevolved Partner



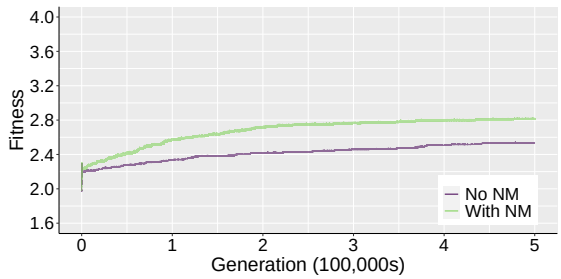
(b) One Environment with Random Partners



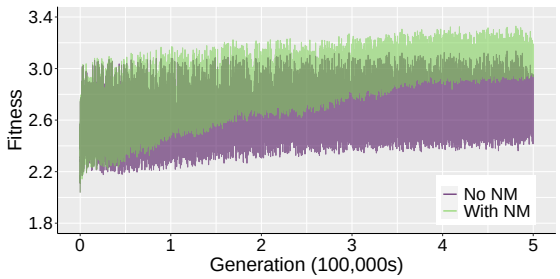
(c) Two Environments with a Coevolved Partner



(d) Two Environments with Random Partners



(e) Four Environments with Coevolved Partners



(f) Four Environments with Random Partners

Figure 6.1: The mean best-in-population fitnesses of agents in the RCD evolving for 500,000 generations, with and without neuromodulation (NM). At each generation, agents are evaluated on: one ((a) and (b)), two ((c) and (d), or four ((e) and (f)) environments, with coevolved ((a), (c) and (e)) or random ((b), (d) and (f)) partners. For (a) and (b), agents share one environment so the maximum fitness is 1.0. For (c) and (d), agents evolve alone, then in a shared environment; the maximum fitness is 1.7. For (e) and (f), agents evolve in an environment: alone, shared, alone, then shared; the maximum fitness is 3.4. A fitness of: 0.7 indicates the goal is achieved individually; 0.9 is cooperation; 1.0 is exploitation; ≥ 0.7 is a goal-achieving fitness (Equation 3.1). *Note: y-axis scales between Figures 6.1a and 6.1b, 6.1c and 6.1d, and 6.1e and 6.1f are comparable, but are not comparable otherwise as agents are evaluated on different numbers of environments and thus have different maximum fitnesses. Also note the data for (b), (d) and (f) are downsampled by a factor of 10.*

and affect the environment in an unpredictable way, resulting in the agent experiencing more diverse environmental conditions. The environment the agents evolve in is likely to have a higher level of uncertainty and change as a result, compared to an environment in which two agents coevolve with one another.

The mean best-in-population fitness over time of agents evolving with random partners, both with and without neuromodulation, is presented in Figure 6.1b. Only 60% of non-modulatory agents are able to achieve their goal after evolving with different partners at each generation, compared to 86% of modulatory agents (Table 6.2). Behavioural plasticity is therefore observed to be beneficial to agents that learn from and navigate an environment inhabited by changing, unpredictable partners.

When comparing Figures 6.1a and 6.1b, it appears that the predictability of another agent's actions within the environment can dramatically affect the evolution of, and the fitness received by, agents. The actions of random partners create high levels of variability, and as such, the fluctuations in fitness in each individual run cause the overall mean fitness to fluctuate by large amounts too. This is because each random partner can potentially change the environment in a different way, meaning agents may not necessarily know how to achieve their goals at each generation; previously successful behaviour may not be successful in the future due to the actions of other agents. This can be seen in Figure 6.1b, which is in fact a line graph, but is erratic since the mean changes by large amounts between generations. It does however seem that the magnitude of these fluctuations decreases over time, which indicates experiencing more variability helps agents to respond appropriately to environmental changes to achieve their goals. This observation can also be seen more obviously in modulatory agents than non-modulatory agents, demonstrating that behavioural plasticity can help agents to succeed in uncertain and environments. Despite the large number of fluctuations in fitness during evolution when agents evolve with random partners, the mean best-in-population fitness is higher earlier on in evolution than when agents coevolve. Interestingly, this is not a result of more agents achieving their goal when evolving with random partners compared to coevolved partners – in fact, the opposite is true when looking at the breakdown of goal-achieving agents presented in Table 6.2. There is however a shift from agents that tend to exploit the other or achieve their goals independently when coevolving, to those that are more inclined to cooperate with random partners.

Agents may evolve to rely on an agent it coevolves with to achieve its goals, thus exploiting an environment that has become predictable in its nature. As agents do not possess the capacity to perceive, learn or reason about other agents that may or may not exist within the environment, exploitation or 'relying' on the actions of others can not be intentional, nor can it be understood; from the agent's perspective, the actions of the other agent within the environment just appear as environmental changes that – if predictable enough – may be taken advantage of without any understanding of the cause, or of the implications of such an exploitation. Exploitative behaviour however carries a higher risk when partners are random, as not all partners may perform behaviour that can be exploited; this would leave the agent unable to achieve its goal. Evolving with random partners seems to incentivise agents to evolve goal-achieving behaviour that involves the other agent in the environment somehow – either through cooperation or exploitation. This is an interesting observation, as the agents have no capacity to learn of or perceive others within the environment.

6.4.3 Evolving in Two Environments with a Coevolved Partner

The following sets of experiments in this study explore how agent evolution and goal-achievement can be affected when another agent may or may not exist in the environment – the environment can therefore vary in each evaluation. Here, agents are evaluated on two independent RCD instances at each generation – the first in which the agent exists alone, and coevolves with a partner agent in the second. The ability of an agent to achieve goals when faced with different environmental conditions is therefore tested, as behaviour that is successful in one environment or context may not be as successful in another. Agents thus experience a higher level of environmental variability than previous experiments due to evolving in both single- and multi-agent environments.

Figure 6.1c presents the mean best-in-population fitness over time for both modulatory and non-modulatory agents. Similarly to the results presented in Section 6.4.1, neuromodulation appears to be beneficial to agents both immediately and throughout evolution, as the mean best-in-population fitness is higher in modulatory agents than in non-modulatory agents. Further, 37% of agents are able to achieve their goal in both environments without neuromodulation, which increases to 66% in modulatory agents (Table 6.3). This benefit is also seen when looking at the number of successful agents in each of the two environments

Partner	NM	Env	Fitness (% of Runs)				Goals Achieved			
			0.7 (Indep)	0.9 (Coop)	1.0 (Exploit)	≥ 0.7 (GAF)	0	1	2	$\bar{x}$
Coevolved	No	1	37	0	0	37	35	28	37	1.02
		2	37	2	26	65				
	Yes	1	66	0	0	66	4	30	66	1.62
		2	66	0	30	96				
Random	No	1	33	0	0	33	65	7	28	0.63
		2	5	25	0	30				
	Yes	1	77	0	0	77	23	10	67	1.44
		2	10	57	0	67				

Table 6.3: The percentage of modulatory and non-modulatory agents across 100 runs that achieved their goal with common fitnesses in each environment (Env) after evolution. Agents are evaluated alone, then with coevolved or random partners. A fitness of 0.7 is achieving the goal independently by building a bridge with two Stones (Indep); 0.9 is cooperation (Coop) by sharing the cost of bridge-building; 1.0 is exploitation (Exploit); ≥ 0.7 is a goal-achieving fitness (GAF).

individually. For the first environment, 37% of non-modulatory agents achieve their goal compared to 66% of modulatory agents, whereas in the second this rises to 65% and 96% respectively. Both modulatory and non-modulatory agents have a higher success rate when there is another agent in the environment, as agents who cannot achieve the goal when alone can capitalise on the actions of the other agent to receive a higher fitness. Ultimately, behavioural plasticity is shown to have a positive effect on the success rate of agents that both exist alone, and coevolve with another.

6.4.4 Evolving in Two Environments with Random Partners

In this part of the study, agents are first evaluated on an environment in which they exist alone, and are then evaluated on an environment shared with random partners. Agents encounter a higher level of uncertainty and environmental change than in previous experiments, as they are tasked with evolving successful behaviour in both single- and multi-agent environments where the partner agents have the potential to affect the environment in different ways at each generation.

The mean best-in-population fitness over time for both modulatory and non-modulatory agents is presented in Figure 6.1d. Like with the results discussed in Section 6.4.2, the mean best-in-population fitness fluctuates often throughout evolution, with the fluctuations reducing over time; this indicates that evolving behaviour that is successful both when the agents are alone, and when with random partners is a difficult task, but becomes easier

as agents experience more variety of environmental conditions. This effect is seen to be greater in modulatory agents, as the mean best-in-population fitness is greater at the end of evolution compared to non-modulatory agents. To evidence this further, only 28% of non-modulatory agents are able to achieve their goal in both environments after evolution, compared to 67% of modulatory agents (Table 6.3). In the first environment where agents exist alone, 33% of non-modulatory agents were successful compared to 77% of modulatory agents. In the second environment that is shared with random partners, only 30% of non-modulatory agents were successful compared to 67% of modulatory agents. For each of the two individual environments, more modulatory agents were successful after evolution than non-modulatory agents; additionally, more agents overall were successful in the first environment than the second environment, which provides further evidence that evolving successful behaviour that generalises across multiple environments and multiple partners is a difficult task.

Similarly to the results observed in the first two sets of experiments, agents that evolve with random partners when evaluated on two environments at each generation achieve a higher mean best-in-population fitness than those that coevolve; evolution is however much more volatile as a result of the agents attempting to achieve their goals in unforeseen environmental conditions (Figures 6.1c and 6.1d). Extending on from the observations discussed in Section 6.4.2, Table 6.3 shows a distinct divergence in evolved behaviour between agents that evolve with either random or coevolved partners. When the environment has an element of predictability arising from coevolving with a single partner, agents almost exclusively evolve to achieve their goals either independently, or by exploiting the other agent within the environment; the same percentage of agents evolve to achieve their goals independently in both environments in this case. This demonstrates that exploitative behaviour may be beneficial when the presence of another agent can be relied on (or more specifically, the environmental changes in which that agent creates, as agents cannot perceive or reason about one another or the results of their actions), but is not helpful when agents find themselves in a situation in which they must fend for themselves. In contrast, agents that have experienced a wide range of environmental conditions from evolving with random partners are more likely to cooperate with their partner agents than to independently achieve their goal – with none opting to exploit their partner. Consequently, these agents that evolve with random partners appear to be capable of altering the way in which they interact with the

environment to achieve their goals; they can take actions to increase their fitness without then negatively affecting their ability to achieve their goals when existing in an environment alone.

6.4.5 Evolving in Four Environments with Coevolved Partners

The results thus far show that the variability experienced during evolution, and therefore the range of environmental conditions encountered as a result of the actions of others, can affect an agent's ability to achieve its goals. In these final two sets of experiments, the number of environments that agents are evaluated on at each generation is increased to four, thus exploring how experiencing a high levels of variability affects evolution and goal-achievement; agents are evaluated in an environment alone, with a partner agent A, alone for a second time, and finally with a partner agent B. As there is more variability across all four environments at each generation than with one or two environments, it is hypothesised that evolving successful solutions in these experiments will therefore be more difficult.

Figure 6.1e shows the mean best-in-population fitness over time of agents that are evaluated on four environments at each generation, with coevolved partners, both with and without neuromodulation. In line with the results discussed in Sections 6.4.1 and 6.4.3, modulatory agents appear to receive an immediate and sustained benefit compared to non-modulatory agents, and outperform their non-modulatory counterparts throughout evolution. 75% of non-modulatory agents were successful in at least one environment, however only 26% were able to achieve their goal in each of the four environments at the end of evolution (Table 6.4). In comparison, 98% of modulatory agents were successful in at least one environment, but only 66% were successful in all four. This is an indication that evolving goal-achieving behaviour that generalises across four different environments is a difficult task, but behavioural plasticity can help agents to overcome this difficulty. To break this down further, 26% of non-modulatory agents were successful in each of the single-agent environments, and 62% and 57% were successful in the second and fourth environments respectively; in comparison, 66% of modulatory agents were successful in both of the single-agent environments, and 92% and 96% were successful in second and fourth environments respectively. Neuromodulation therefore clearly provides a benefit to agents that are tasked with achieving goals in many environments that vary from one to another, and where another agent may or may not exist within said environments. Modulatory agents

Partners	NM	Env	Fitness (% of Runs)				Goals Achieved					
			0.7 (Indep)	0.9 (Coop)	1.0 (Exploit)	≥ 0.7 (GAF)	0	1	2	3	4	$\bar{x}$
Coevolved	No	1	26	0	0	26	25	31	18	0	26	1.71
		2	26	3	33	62						
		3	26	0	0	26						
		4	26	3	28	57						
	Yes	1	66	0	0	66	2	8	24	0	66	3.2
		2	65	1	26	92						
		3	66	0	0	66						
		4	66	0	30	96						
Random	No	1	31	0	0	31	64	2	3	4	27	1.28
		2	7	24	0	31						
		3	32	0	0	32						
		4	21	13	0	34						
	Yes	1	86	0	0	86	13	1	1	6	79	3.37
		2	12	68	0	80						
		3	85	0	0	85						
		4	63	23	0	86						

Table 6.4: The percentage of modulatory and non-modulatory agents across 100 runs that achieved their goal with common fitnesses in each environment (Env) after evolution. Agents are evaluated alone, with a partner, alone, then with a partner. Partners are coevolved or random. A fitness of 0.7 is achieving the goal independently by building a bridge with two Stones (Indep); 0.9 is cooperation (Coop) by sharing the cost of bridge-building; 1.0 is exploitation (Exploit); ≥ 0.7 is a goal-achieving fitness (GAF).

can therefore be expected to achieve their goal in more environments than non-modulatory agents, and are also more likely to succeed in all four environments.

6.4.6 Evolving in Four Environments with Random Partners

In these last experiments, agents are evaluated on four environments at each generation with random partners. The variability of these environments therefore increases in comparison to previous experiments, as agents are exposed to many different situations and scenarios in which they can learn from, due to the actions of unknown, unpredictable partners within the environment itself.

The mean best-in-population fitness over time of both non-modulatory and modulatory agents evaluated on four environments at each generation, with random partners, is shown in Figure 6.1f. Similarly to the results discussed in Sections 6.4.2 and 6.4.4, the uncertainty that arises as a result of the unpredictable actions of other agents within the environment appears to increase the fluctuations in the mean best-in-population fitness throughout evolution for all agents. As evolution progresses, these fluctuations reduce – however this is more obvious in modulatory agents than non-modulatory agents. This shows that behavioural

plasticity can positively affect evolution and the ability to achieve goals in highly variable and uncertain environments.

At the end of evolution, only 36% of non-modulatory agents were successful in at least one environment, and 27% were successful in all four environments. This is compared to 87% of modulatory agents that were successful in at least one environment, and 79% in all four (Table 6.4). Modulatory agents are also more successful than their non-modulatory counterparts in each individual environment: 31%, 31%, 32% and 34% of non-modulatory agents were successful in environments one, two, three and four respectively, compared to 86%, 80%, 85% and 86% of modulatory agents in the same environments. Not only are modulatory agents more likely to achieve their goal in each individual environment, but they are also more likely to have evolved behaviour that enables them to achieve their goal in all four environments. Neuromodulation is therefore observed to have a positive effect on agents when they evolve in environments with high levels of uncertainty and variability, enabling them to achieve success more often than those that do not use neuromodulation.

A similar trend is observed in agents that are evaluated on four environments at each generation, compared to just one or two environments; the fitness that agents receive during evolution fluctuates more when there is more environmental uncertainty, but agents are able to achieve a higher mean best-in-population fitness more quickly than those that experience more certainty within the environment (Figures 6.1e and 6.1f). Whilst fewer agents are able to achieve their goal in at least one environment when evolving with random partners compared to those that coevolve, *more* agents are actually able to achieve their goals in all four environments when faced with higher levels of environmental uncertainty (Table 6.4). The qualitative shift in goal-achieving behaviour described throughout this chapter is more pronounced in these experiments, due to the increased difficulty and variability involved in completing the task. Specifically, exploitative behaviour is more prevalent when the actions of unknown, coevolved partners cause the environment to become more predictable; this exploitative behaviour is not useful when agents exist alone, showing that no beneficial behaviour is evolved. However, when environmental uncertainty arises through evolving with random partners, cooperative behaviour is favoured. Agents are shown to possess the capacity to achieve their goals individually as they can achieve their goals when existing alone in an environment; they are also shown to have evolved the ability to take actions to increase their fitness whenever possible, through cooperating with other agents uninten-

tionally. This shows that agents who have experienced and learnt from a wide range of environmental conditions – by both evolving alone, and attempting to achieve their goals in an environment that can be changed in unpredictable ways by unknown partners – are likely to evolve behaviour that enables them to make the most of the situation they find themselves in, without that affecting their ability to succeed when environmental conditions inevitably change.

6.5 Analysing the Effects of Plasticity and Variability on Agent Evolution

Thus far, the mean best-in-population fitnesses of agents in each experiment are discussed; in this section, a comprehensive analysis of agent evolution is presented, in order to understand how the level of environmental uncertainty caused by either the actions or presence of others can affect behaviour and the ability to achieve goals. Further, the effect that behavioural plasticity can have on agents in these different environments is also analysed, to ascertain the extent to which neuromodulation affects both an agent's ability to achieve goals and the evolutionary process itself.

6.5.1 Analysing Fitness

As the number of environments that agents are evaluated on at each generation differs between each experiment in this study, the maximum achievable fitness at each generation increases with the number of environmental evaluations. Agents evolving in one, two or four environments have a theoretical maximum fitness of 1.0, 2.0, or 4.0 respectively; as the minimum goal-achieving fitness is 0.7 and a maximum fitness of 1.0 may be achieved if an agent exploits the actions of another, the actual range of goal-achieving fitnesses across all environments in each experiment are $0.7 - 1.0$, $1.4 - 1.7$ or $2.8 - 3.4$ respectively[2]. The term 'fitness' is used throughout this section to refer to the best-in-population agent fitness across all 100 runs in an experiment at the end of evolution.

Experiment	Partner	NM	Moment				Median
			Mean	Variance	Skewness	Kurtosis	
1	Coevolved	No	0.754	0.0447	0.0245*	1.36*	0.700
		Yes	0.838*	0.0286 *	−0.282	1.60	0.850*
	Random	No	0.765	0.0500	−0.270 *	1.21*	0.900
		Yes	0.862*	0.0248 *	−1.59	4.17	0.900
2	Coevolved	No	1.29	0.0459	−0.509 *	1.42*	1.40
		Yes	1.42 *	0.009 64*	−2.83	13.4	1.40
	Random	No	1.31	0.0764	−0.163 *	1.21*	1.40
		Yes	1.50 *	0.0546 *	−1.52	3.64	1.60 *
3	Coevolved	No	2.53	0.125	−0.402 *	1.88*	2.50
		Yes	2.82 *	0.0313 *	−2.14	10.4	2.80 *
	Random	No	2.71	0.0818	0.693 *	1.67*	2.50
		Yes	3.05 *	0.0494 *	−1.68	4.72	3.10 *

Table 6.5: Statistical moments and median (to 3 S.F.) of the best-in-population fitness after 500,000 generations; agents are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved or random partners. The lowest mean, variance, skewness, kurtosis, and median for agents with or without neuromodulation (NM) are indicated with an asterisk (*).

Analysing the Fitness Distributions

The statistical moments and the median of the fitness distributions of agents after evolving in the RCD are presented in Table 6.5; further, the underlying distribution of fitnesses for each experiment are visualised as kernel density estimation plots, presented in Figure 6.2. At the end of evolution in all experiments, agents capable of behavioural plasticity have a higher mean fitness, and either the same or higher median fitness than those that are not, both when evolving with coevolved or random partners. Additionally, the variance in fitness is lower in modulatory agents than non-modulatory agents, demonstrating that neuromodulation enables agents to achieve higher fitnesses more consistently than non-modulatory agents – and more often (Tables 6.2, 6.3 and 6.4).

Each fitness distribution for agents evolving with neuromodulation is negatively skewed, indicating that modulatory agents are likely to achieve a higher-than-average fitness; this can be seen in Figure 6.2, as the peaks in the kernel density estimation plots that visualise the fitness distributions of modulatory agents are towards the higher-end of the x-axis. The fitness distributions for non-modulatory agents in each experiment however vary from positively to negatively skewed; saying this, in each experiment, the fitness distributions

[2]When agents are evaluated on either two or four environments, the actual maximum fitness is lower than the theoretical maximum fitness as agents exist in some of the environments alone, and thus can only achieve a maximum fitness of 0.7.

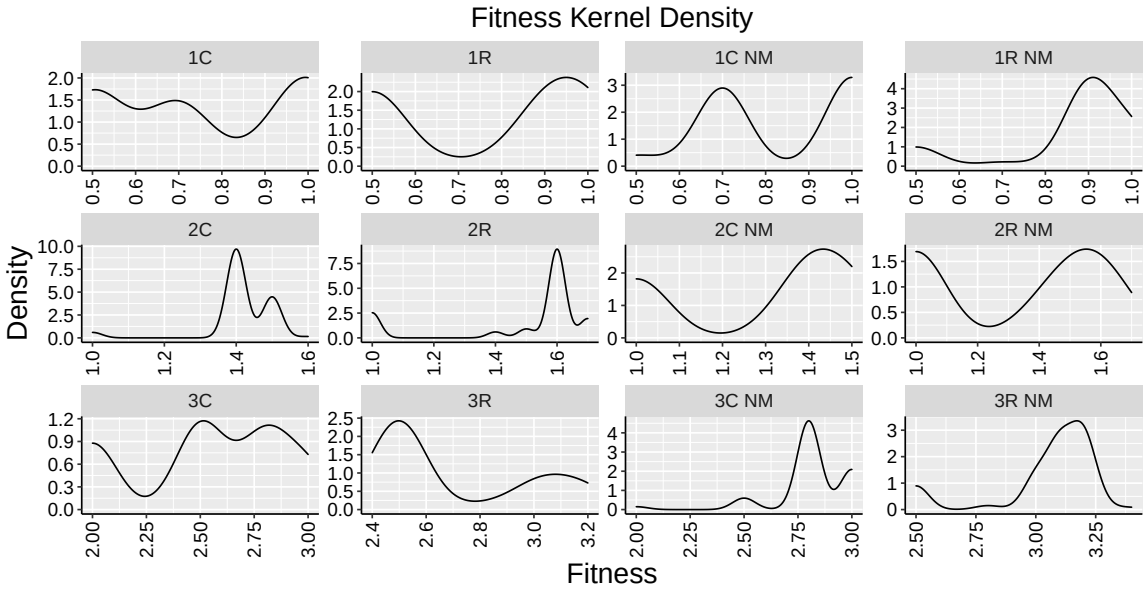


Figure 6.2: Kernel density estimations of the best-in-population fitness of agents that are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved (C) or random (R) partners. Experiments are repeated for non-modulatory and modulatory (NM) agents. *Note: the graph scales are not comparable due to the disparity in the densities between each distribution.*

of non-modulatory agents are less skewed than those of modulatory agents. This indicates that non-modulatory agents are more likely to achieve a fitness around the average whereas modulatory agents are more likely to receive a higher-than-average fitness. In terms of the shape of the distributions, the fitness distributions of modulatory agents are generally leptokurtic, meaning that the kurtosis is greater than that of a normal distribution (which has a kurtosis of 3) [65]; this also indicates more outliers are present at the tails of the distributions. In comparison, all fitness distributions of non-modulatory agents are platykurtic, as the kurtosis is less than 3 in each; the distributions are therefore flatter than that of modulatory agents, and outliers are less frequent as tails are shorter. Overall, these findings indicate that modulatory agents are more likely to achieve a higher fitness than their non-modulatory counterparts, and therefore are more likely to achieve their goal. This is seen in all experiments, regardless of how many environments that agents are evaluated on at each generation.

Statistical Tests – Comparing Non-Modulatory and Modulatory Agents

Statistical tests were carried out to conclude whether there is any significant difference between the best-in-population fitnesses achieved after evolution by modulatory and non-

Exp	Partner	Statistical Test Alternative Hypothesis			z	r
		$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$		
1	Coevolved	1.594×10^{-2} *	7.970×10^{-3} *	9.922×10^{-1}	−2.413	−0.2413 (S)
1	Random	1.871×10^{-3} *	9.355×10^{-4} *	9.991×10^{-1}	−3.112	−0.3112 (M)
2	Coevolved	2.331×10^{-6} *	1.165×10^{-6} *	1	−4.725	−0.4725 (M)
2	Random	1.248×10^{-6} *	6.239×10^{-7} *	1	−4.851	−0.4851 (M)
3	Coevolved	4.243×10^{-9} *	2.121×10^{-9} *	1	−5.877	−0.5877 (L)
3	Random	1.951×10^{-11}*	9.757×10^{-12}*	1	−6.712	−0.6712 (L)

Table 6.6: Wilcoxon Signed Rank statistical tests comparing the best-in-population fitnesses of non-modulatory (m_n) and modulatory (m_m) agents after 500,000 generations; agents are evaluated on: (1) one, (2) two or (3) four environments in each experiment (Exp). p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

modulatory agents in each experiment. Firstly, as the fitness distributions described in Table 6.5 are skewed to some degree and have either high or low kurtosis, a Shapiro-Wilk test for normality was conducted on the results of each experiment; this test was chosen specifically because it is powerful for a range of distributions [241]. The results of these tests were significant at the $p < 0.05$ level for each distribution, indicating that the distributions are non-normal; these results can be found in Table B.7 in Appendix B.3. Consequently, one two-tailed and two one-tailed Wilcoxon Signed Rank statistical tests were conducted for each experiment to ascertain whether behavioural plasticity has any effect on the best-in-population fitness of agents after evolution; the results of these tests are presented in Table 6.6. There is a significant difference in the median best-in-population agent fitness after evolution when non-modulatory (m_n) and modulatory (m_m) agents in each experiment are compared ($p < 0.05$, $m_n \neq m_m$); further, the median fitness of non-modulatory agents is observed to be significantly lower than modulatory agents in each experiment ($p < 0.05$, $m_n < m_m$). Behavioural plasticity is therefore observed to have a positive impact on the fitness that agents receive at the end of evolution – even when they are evaluated on an increasing number of environments at each generation.

To analyse the effect that behavioural plasticity has on agent fitness further, the effect size estimate r can be used to measure the magnitude of such an effect (Section 3.3), presented in Table 6.6. As the effect size is negative for each experiment, the fitness received by modulatory agents is higher than their non-modulatory counterparts; this effect is also stronger when agents evolve with random partners, in addition to when the number of environments agents are evaluated on at each generation is increased. While the statistical tests

discussed previously indicate that modulatory agents achieve significantly higher fitnesses than non-modulatory agents in each experiment ($p < 0.05$), behavioural plasticity is shown to have a smaller effect on agents that evolve in less variable environments. Concretely, the least variable environment by this definition is that which evaluates agents on one environment, and evolves them with a single, coevolved partner; in this case, $r = -0.2413$, which is deemed a 'small' effect using the guidelines outlined by Cohen [58]. In comparison, behavioural plasticity has a 'large' effect on fitness ($r = -0.6712$) in the most variable experiment, which evaluates agents on four environments and with random partners. This is evidence of a correlation between the benefit of behavioural plasticity and the variability of the environment, where the benefit of neuromodulation increases with environmental variability. Upon closer inspection, this benefit is felt more strongly by agents that evolve with random partners compared to those that evolve with coevolved partners, as the actions of random partners increase environmental variability further.

As discussed in Section 6.2.1, behavioural plasticity can regulate behaviour by temporarily changing synaptic activity locally, or learning if synaptic strengths are modulated over time [196] – the former of which is explored in this chapter. Behavioural plasticity via neuromodulation is beneficial to agents in variable environments, because agents are able to change their phenotype (their behaviour) within their lifetime, without affecting their genotype (deliberative network); any behavioural changes will be short-lived. This would explain why plasticity is more beneficial as variability increases, as the ability to reversibly and temporarily change behaviour depending on the environment would happen more often when the environment differs at each generation.

Statistical Tests – Comparing Evolution with Coevolved and Random Partners

In addition to the analysis of the fitnesses that non-modulatory and modulatory agents achieve after evolution, here the fitness that agents receive when evolving with either coevolved (p_c) or random (p_r) partners is also compared. This is to explore whether the increase in environmental variability, caused by the unpredictable actions of random partners, has an effect on agent fitness. As above, Wilcoxon Signed Rank statistical tests were chosen as a suitable test, as each distribution was ascertained to be non-normal by a significant Shapiro-Wilk test result ($p < 0.05$, Table B.7 in Appendix B.3). The results of the Wilcoxon Signed Rank tests can be found in Table 6.7. No significant difference was found

Exp	NM	Statistical Test Alternative Hypothesis			z	r	
		$p_c \neq p_r$	$p_c < p_r$	$p_c > p_r$			
1	No	5.632×10^{-1}	2.816×10^{-1}	7.202×10^{-1}	-0.5807	$-0.058\,07$	
1	Yes	3.469×10^{-2} *	1.734×10^{-2} *	9.828×10^{-1}	-2.114	-0.2114	(S)
2	No	3.128×10^{-1}	1.564×10^{-1}	8.448×10^{-1}	-1.012	-0.1012	(S)
2	Yes	1.938×10^{-3} *	9.688×10^{-4} *	9.990×10^{-1}	-3.102	-0.3102	(M)
3	No	2.339×10^{-4} *	1.169×10^{-4} *	9.999×10^{-1}	-3.682	-0.3682	(M)
3	Yes	1.146×10^{-11}*	5.730×10^{-12}*	1	-6.789	-0.6789	(L)

Table 6.7: Wilcoxon Signed Rank statistical tests comparing the best-in-population fitnesses of agents evolving with coevolved (p_c) or random (p_r) partners after 500,000 generations, when they do and do not use neuromodulation (NM); agents are evaluated on: (1) one, (2) two or (3) four environments in each experiment (Exp). p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

between the median fitnesses after evolution of non-modulatory agents that evolve with either coevolved or random partners, when they are evaluated on one or two environments at each generation. However, the corresponding modulatory approaches are found to have a lower median fitness when evolving with coevolved partners, than with random partners in the same experiments ($p < 0.05$, Table 6.7). Further to this, when agents are evaluated on four environments at each generation, agents that evolve with coevolved partners receive a significantly lower fitness to those that evolve with random partners – regardless of whether they exhibit behavioural plasticity or not ($p < 0.05$). Modulatory agents are therefore observed to receive a higher fitness at the end of evolution when they evolve with random partners than when the partner is coevolved – a finding which is corroborated with the data presented in Table 6.5.

To analyse these statements more concretely, the effect size estimate r (introduced in Section 3.6.2) can be considered, which captures the magnitude of how much the type of partner (coevolved or random) affects an agent's received fitness; in doing this, the type of partner is shown to have a larger effect on agents that exhibit behavioural plasticity than those that do not (Table 6.7). Specifically, as the effect size estimate r for each experiment is negative, agents that evolve with random partners are observed to receive a higher fitness than when they evolve with coevolved partners; this effect is stronger in modulatory agents than non-modulatory agents, demonstrating that behavioural plasticity via neuromodulation can help agents to overcome the environmental uncertainty caused by the unpredictable actions of others. In other words, agents that experience a wide range of environmental conditions as a consequence of evolving with random partners – especially

when the agent can exhibit behavioural plasticity – are able to achieve higher fitnesses than agents that evolve with more certainty within the environment from evolving with a coevolved partner; this effect is stronger as the diversity of experiences and uncertainty within the environment increase.

Further, the type of partner has a larger effect on the fitness received by agents as the number of environments they are evaluated on at each generation increases. For example, $r = -0.05807$ when non-modulatory agents are evaluated on one environment, but $r = -0.3682$ when non-modulatory agents are evaluated on four environments; using the guidelines outlined by Cohen [58], the type of partner (coevolved or random) has a negligible ($r < 0.1$, which is less than the threshold that Cohen defines as a 'small' effect) or medium effect on agent fitness respectively, with the fitness being higher when partners differ at each generation. This is similarly observed in modulatory agents, however the effect size is larger overall; the type of partner has a small effect ($r = -0.2114$) when agents are evaluated on one environment, but a large effect ($r = -0.6789$) when evaluated on four environments. The divergence in fitnesses received by agents that evolve with coevolved partners, and those that evolve with random partners, therefore increases as the number of environments that agents are evaluated on increases. This shows that evolving in an environment where it is possible to experience a diverse range of scenarios can help agents to evolve diverse behaviours, enabling them to achieve their goals when the environment is uncertain and unpredictable; behavioural plasticity also provides an additional benefit by allowing agents to temporarily alter their behaviour in response to changing environmental stimuli.

These results demonstrate that the predictability of the actions of others within the environment can impact goal-achieving behaviour in agents during evolution, and the fitness received as a consequence at the end of evolution. Evolving in an environment with random partners is shown to increase the fitness agents receive at the end of evolution; this effect is stronger both when agents exhibit behavioural plasticity, and when agents are evaluated on more environments at each generation. This suggests that evolution is able to find higher fitness goal-achieving solutions when agents are exposed to more diverse experiences; consequently, agents that have encountered many different environmental conditions as a result of evolving with random partners are observed to evolve goal-achieving behaviour that is more reliable and likely to contribute to the achievement of the agent's goals – even when the presence and actions of others in the environment is uncertain.

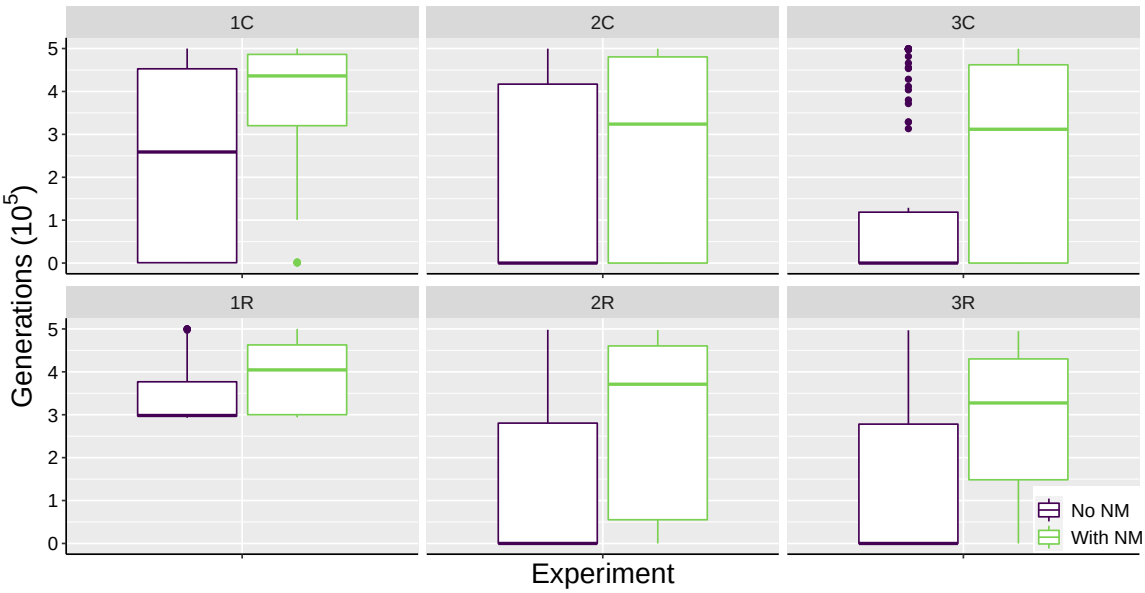


Figure 6.3: Box plots depicting the number of generations agents receive a goal-achieving fitness (≥ 0.7) on all environments they are evaluated on at each generation, during 500,000 generations of evolution, both with and without neuromodulation; agents are evaluated on: (1) one, (2) two or (3) four environments in each experiment, with coevolved (C) or random (R) partners.

6.5.2 Analysing Goal-Achievement Over Evolution

Behavioural plasticity has been shown to increase the expected fitness of agents that evolve in environments with either coevolved or random partners, when they are evaluated on an increasing number of environments at each generation; neuromodulation is also shown to increase the likelihood that an agent will achieve its goals compared to agents that are not capable of behavioural plasticity.

In this section, the number of generations that agents receive a goal-achieving fitness (≥ 0.7) in each environment they are evaluated on, over the course of 500,000 generations of evolution, are compared to analyse how consistently agents are able to achieve goals when they either do or do not use neuromodulation; this is shown in the form of box plots in Figure 6.3, with the detailed statistics presented in Table 6.8. In each experiment, the first, second and third quartiles are either the same or higher in modulatory agents than non-modulatory agents, indicating that the number of generations that modulatory agents achieve their goals is higher than in non-modulatory agents. Behavioural plasticity is therefore observed to have a positive effect on how often agents are able to achieve their goals throughout evolution; this stands true both when the agents evolve in environments that are unpredictable as a result of the actions of random partners, as well as when the

Experiment	Partner	NM	Min	Q_1	Median	Mean	Q_3	Max
1	Coevolved	No	16	852	258 916	244 726	452 762	499 992
		Yes	399	320 212	436 195	385 511	486 331	499 977
	Random	No	292 960	296 994	298 262	343 336	376 802	499 921
		Yes	293 851	300 090	404 514	397 230	462 697	499 951
2	Coevolved	No	0	0	6	153 879	417 033	499 781
		Yes	0	22	323 905	258 679	480 500	499 518
	Random	No	0	0	2	126 750	280 481	498 001
		Yes	0	55 278	371 087	289 040	460 377	497 546
3	Coevolved	No	0	0	0	106 781	118 640	499 536
		Yes	0	8	311 899	254 337	462 123	499 491
	Random	No	0	0	0	117 396	278 120	496 856
		Yes	0	148 452	327 471	281 954	430 108	494 884

Table 6.8: The minimum, maximum number of generations that agents achieve their goal, along with the first, second and third quartiles (Q_1, median, and Q_3 respectively), and the mean. Agents evolve for 500,000 generations and are evaluated on (1) one, (2) two, or (3) four environments, with or without neuromodulation (NM). Values are rounded to the nearest whole number.

difficulty of the task – through the number of environments that agents are evaluated on – increases.

Statistical tests were then conducted to analyse whether any significant difference exists between the non-modulatory (m_n) and modulatory (m_m) approaches, in terms of the number of 'successful' generations during evolution; a 'successful' generation is one in which an agent receives a goal-achieving fitness of ≥ 0.7 in each environment that it is evaluated on at each generation. Firstly, Shapiro-Wilk tests concluded that the distribution for each experiment was non-normal, as the results were significant at $p < 0.05$; these results can be found in Table B.8 in Appendix B.3. Wilcoxon Signed Rank statistical tests were chosen as an appropriate statistical test, in order to compare the medians of each distribution – the results of which are presented in Table 6.9. The median number of generations non-modulatory and modulatory agents are successful in each experiment were ascertained to be unequal in each experiment ($p < 0.05$, $m_n \neq m_m$), with non-modulatory agents achieving their goals for a smaller proportion of evolution when compared to modulatory agents ($p < 0.05$, $m_n < m_m$). These results provide further evidence of the positive impact that behavioural plasticity has on agent evolution when the environment is unpredictable or challenging; this benefit of using neuromodulation is seen when agents are paired with either coevolved or random partners, as well as in experiments where the agents are evaluated on more than one environment at each generation. This shows that modulatory agents are able to achieve

Experiment	Partner	Statistical Test Alternative Hypothesis			z	r
		$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$		
1	Coevolved	2.177×10^{-6} *	1.088×10^{-6} *	1	−4.738	−0.4738 (M)
1	Random	6.193×10^{-6} *	3.097×10^{-6} *	1	−4.521	−0.4521 (M)
2	Coevolved	2.331×10^{-6} *	1.165×10^{-6} *	1	−2.640	−0.2640 (S)
2	Random	1.248×10^{-6} *	6.239×10^{-7} *	1	−4.850	−0.4850 (M)
3	Coevolved	4.243×10^{-9} *	2.121×10^{-9} *	1	−4.624	−0.4624 (M)
3	Random	1.951×10^{-11}*	9.757×10^{-12}*	1	−5.450	−0.5450 (L)

Table 6.9: Wilcoxon Signed Rank statistical tests comparing the number of generations that the best-in-population non-modulatory (m_n) and modulatory (m_m) agents receive a goal-achieving fitness (≥ 0.7) in each experiment; agents are evaluated on: (1) one, (2) two or (3) four environments, with coevolved or random partners. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

their goals more often during evolution than non-modulatory agents.

Agents capable of behavioural plasticity are shown to achieve their goals for more generations than non-modulatory agents in each experiment, indicated by a negative effect size estimate r (Table 6.9). This effect size tends to be stronger in agents that evolve with random partners than those that evolve with coevolved partners; the exception to this is that the effect size of behavioural plasticity on agents that are evaluated on one environment is marginally smaller in agents that evolve with random partners to coevolved partners ($r = -0.4521$ and $r = -0.4738$ respectively), with both still indicating that behavioural plasticity has a 'medium' effect on the agent success rate [58]. For agents that evolve with random partners, the effect size increases with the number of environments they are evaluated on at each generation, but such a trend is not as clear in agents that instead evolve with coevolved partners. A note worth making is that while the r value for agents that are evaluated on two environments with coevolved partners indicates that behavioural plasticity has a small effect on the number of successful generations during evolution, it is however noticeably lower than that of the other experiments. This difference could be explained by the spread of the non-modulatory and modulatory distributions presented in Figure 6.3 (plot 2C), which are more similar than observed in the other experiments.

6.5.3 Analysing the Effect of Behavioural Plasticity on Volatility

Thus far, agents capable of behavioural plasticity are shown to receive higher fitnesses after evolution, and more often during evolution, than those that are not. The analyses presented in Chapter 5 ascertain that neuromodulation enables agents to receive higher fitnesses when

evolving to solve single- and multi-stage tasks in single- and multi-agent environments – but at the cost of evolutionary volatility. Here, further analysis is conducted to discover whether this same trade-off is present when agents evolve in environments with increasing variability, through either the predictability of the partner agent or the number of environments that agents are evaluated on at each generation. The three volatility metrics (SDoT, CACoT and CCoT) presented in Section 3.6.1 are therefore used to analyse and compare the evolutionary volatility experienced by non-modulatory and modulatory agents in the experimental study. The statistical moments and medians for each metric are presented in Table 6.10; for each of the 100 runs of each experiment, a value for each metric was calculated using the best-in-population fitness at each generation across 500,000 generations of evolution.

SDoT – Standard Deviation over Time

Modulatory agents evolving with random partners are generally found to have a lower mean and median SDoT (except when agents are evaluated on four environments with random partners, which could be due to a higher variance), with a higher variance than their non-modulatory counterparts (Table 6.10). However, modulatory agents evolving with coevolved partners experience the opposite, as they have a higher mean and median SDoT, and lower variance than non-modulatory agents. Additionally, agents that evolve with a coevolved partner compared to random partners have a lower mean and median SDoT and higher variance in all experiments.

The SDoT distributions for agents using neuromodulation are less skewed, with generally lower kurtosis, than those not using neuromodulation; the shapes of the distributions can be seen in Figure 6.4, which visualise the kernel density estimation for each experiment. Agents that evolve with coevolved partners generally have positive skewness, meaning SDoT values are likely to be on the lower end of the distribution; agents evolving with random partners however generally are negatively skewed, and have higher SDoT values as a result. The SDoT distributions of non-modulatory agents that evolve with random partners are leptokurtic in each experiment, as kurtosis is greater than 3 (Table 6.10); outliers are therefore more likely. This is especially evident in the case of non-modulatory agents that are evaluated on four environments at each generation, when they evolve with random partners (Figure 6.4); the distribution has a large peak around 0.22, with very long tails that indicate outliers. One possible explanation of why this distribution is disparate from

Metric	Exp	P	NM	Moment				Median
				Mean	Variance	Skewness	Kurtosis	
SDoT	1	C	No	0.0802*	0.007 35	0.763	2.07*	0.0310*
			Yes	0.0981	0.005 94*	0.636 *	2.16	0.0804
		R	No	0.213	0.003 11*	−1.71	4.64	0.245
			Yes	0.176 *	0.003 68	−0.433 *	1.98*	0.189 *
	2	C	No	0.0668*	0.006 92	0.975	2.27	0.0120*
			Yes	0.109	0.005 78*	0.190 *	1.70*	0.107
		R	No	0.207	0.004 62	−1.44	3.30	0.245
			Yes	0.180 *	0.004 61*	−0.510 *	1.81*	0.202 *
	3	C	No	0.130 *	0.0190	0.700	2.02	0.0301*
			Yes	0.203	0.0153 *	−0.001 68*	1.84*	0.206
		R	No	0.227 *	0.004 41*	0.545	4.02	0.219 *
			Yes	0.280	0.008 48	−0.543 *	2.14*	0.299
CACoT	1	C	No	40.8 *	6560 *	5.15	37.2	13.3 *
			Yes	97.1	16 400	3.38 *	19.0 *	46.1
		R	No	12 000	13 100 000 *	−1.26	3.04	14 200
			Yes	9130 *	14 900 000	0.148 *	1.56*	8790 *
	2	C	No	41.6 *	7150 *	6.90	59.4	18.1 *
			Yes	201	48 500	1.52 *	4.69*	97.5
		R	No	11 400	17 900 000	−0.945	2.12	14 200
			Yes	8220 *	17 500 000 *	0.484 *	1.63*	6510 *
	3	C	No	87.3 *	36 600 *	7.55	67.3	42.4 *
			Yes	442	188 000	1.62 *	6.26*	283
		R	No	11 700	4 880 000	−1.13	2.57	13 000
			Yes	9940 *	3 730 000 *	0.434 *	1.89*	9420 *
CCoT	1	C	No	174 *	129 000 *	3.38 *	15.2 *	35 *
			Yes	373	360 000	4.00	24.2	134
		R	No	29 400 *	1 370 000 *	1.65	5.69	28 900 *
			Yes	30 400	2 230 000	−0.384 *	2.90*	30 600
	2	C	No	167 *	192 000 *	6.68	54.2	46 *
			Yes	892	1 220 000	1.76 *	5.47*	400
		R	No	29 300 *	2 830 000 *	1.48	6.31	28 800 *
			Yes	30 900	3 060 000	−0.138 *	1.65*	31 200
	3	C	No	322 *	897 000 *	7.87	71.2	111 *
			Yes	1850	3 610 000	1.05 *	2.88*	998
		R	No	32 300 *	102 000 000	1.13	2.61	26 000 *
			Yes	41 700	90 800 000 *	−0.345 *	1.81*	43 200

Table 6.10: Statistical moments and median (to 3 S.F.) of the SDoT, CACoT and CCoT volatility metrics; agents are evaluated on (1) one, (2) two or (3) four environments in each experiment (Exp), with either coevolved (C) or random (R) partners. The highest mean and median, and lowest variance, skewness, and kurtosis for agents with or without neuromodulation (NM) are indicated with an asterisk (*).

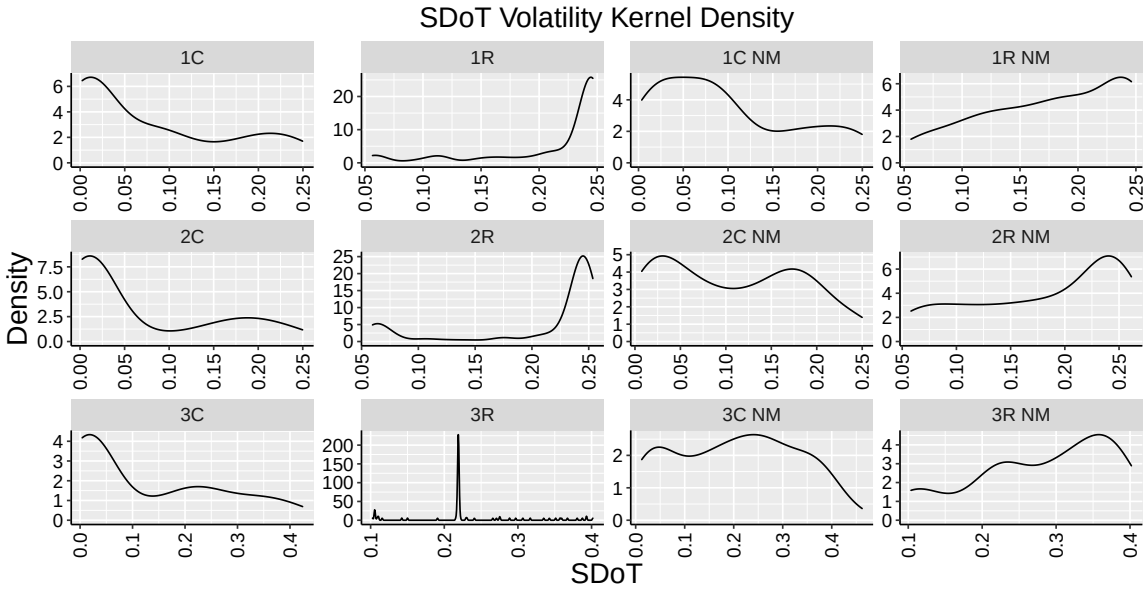


Figure 6.4: Kernel density estimations of the SDoT of agents that are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved (C) or random (R) partners. Experiments are repeated for non-modulatory and modulatory (NM) agents. *Note: the graph scales are not comparable due to the disparity in the densities between each distribution.*

the others could be related to the low numbers of agents that are able to achieve their goals in this experiment (Table 6.4); if an agent cannot retain goal-achieving behaviour that generalises across a number of partners, the fitness will vary across evolution and result in a high SDoT. When non-modulatory agents evolve with random partners, they tend to receive a high SDoT (indicated by the negatively-skewed distributions); a very tall, thin peak in the distribution representing the SDoT of non-modulatory agents evolving with random partners in four environments (plot 3D) thus implies it is predictably difficult for these agents to evolve and maintain goal-achieving behaviour. Whilst non-modulatory agents are observed to have a predictably lower SDoT, modulatory agents may have more variability in fitness during evolution but, they also achieve their goals more often.

When comparing agents that evolve with coevolved or random partners, both modulatory and non-modulatory agents are seen to experience less evolutionary volatility when evolving with coevolved partners than random partners. Evolution appears to be less volatile in modulatory agents that evolve with random partners than non-modulatory agents – as they receive a lower SDoT – but more volatile when they evolve with coevolved partners; this is not necessarily expected because neuromodulation enables agents to temporarily and reversibly change behaviour – and potentially fitness as a result – in

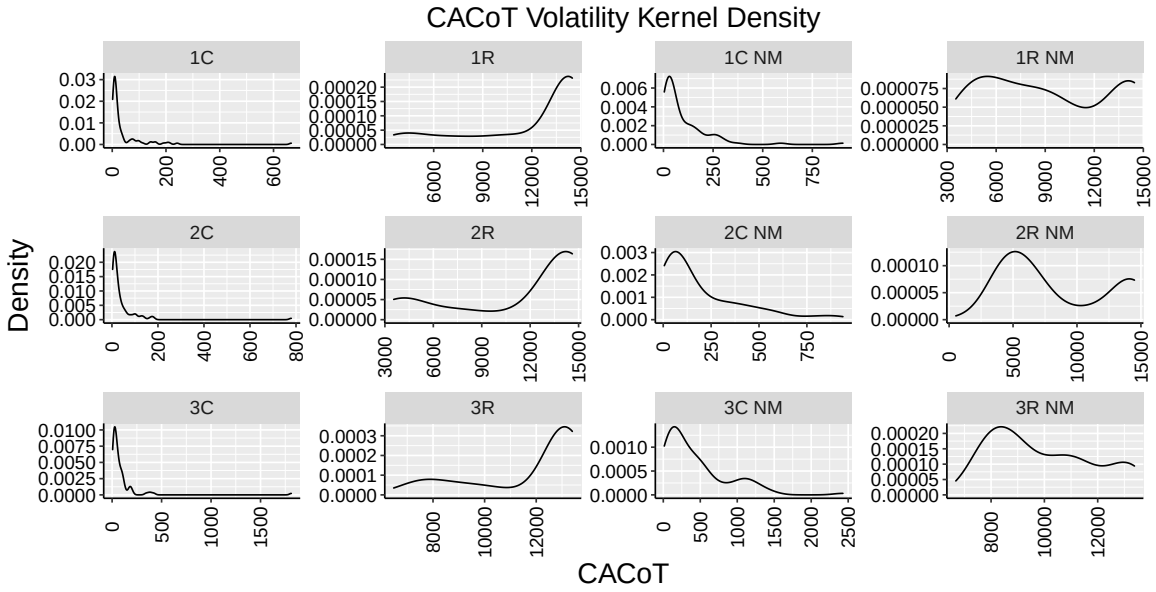


Figure 6.5: Kernel density estimations of the CACoT of agents that are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved (C) or random (R) partners. Experiments are repeated for non-modulatory and modulatory (NM) agents. *Note: the graph scales are not comparable due to the disparity in the densities between each distribution.*

response to environmental changes.

CACoT – Cumulative Absolute Change over Time

When modulatory agents evolve with random partners, they are found to have a lower mean and median CACoT than non-modulatory agents in each experiment (Table 6.10); the opposite is found when modulatory agents evolve with coevolved partners. Further, modulatory and non-modulatory agents that evolve with coevolved partners have a lower mean and median CACoT, with a higher variance compared to when they evolve with random partners. Agents evolving with neuromodulation and coevolved partners are observed to have a higher variance in CACoT than non-modulatory agents, but no other obvious trend regarding the variance was seen in the results. The CACoT distributions of modulatory agents have lower skewness and kurtosis compared to non-modulatory agents; this is the same with agents evolving with random partners compared to coevolved partners. The CACoT distributions of non-modulatory agents evolving with random partners have negative skewness, indicating higher values are more likely. Additionally, the distributions of all agents that evolve with coevolved partners are leptokurtic, as the kurtosis is greater than 3 in each; this is more extreme in non-modulatory agents, indicating outliers are more

likely. This is more evident when looking at the shapes of the kernel density estimation plots in Figure 6.5, as the leptokurtic distributions have large peaks with long tails.

The best-in-population fitness of agents evolving with random partners is therefore observed to fluctuate by larger amounts during evolution compared to those that evolve with a coevolved partner; modulatory agents on the other hand fluctuate less (have a lower CACoT) than non-modulatory agents when evolving with random partners, but the opposite is true when evolving with coevolved partners.

CCoT – Count of Change over Time

When looking at the number of times that an agent's fitness fluctuates during evolution, modulatory agents are generally found to have a higher mean and median CCoT, lower skewness and kurtosis (except when agents are evaluated on one environment with coevolved partners), and a higher variance (except when agents are evaluated on four environments with random partners) than their non-modulatory counterparts (Table 6.10). However, when agents evolve with random partners, the mean, median, and variance in CCoT are higher than when agents evolve with coevolved partners (either with or without neuromodulation), with lower skewness and kurtosis. When looking at the kernel density estimation plots presented in Figure 6.6, the extreme kurtosis and high positive skew (where kurtosis is greater than 3, and skew is greater than 1) observed in non-modulatory agents that evolve with coevolved partners can be seen more obviously due to large peaks and long tails to the right of the plots; while the peaks of these distributions fall around lower CCoT values, outliers that have high CCoT values are to be expected.

Modulatory agents therefore appear to have a best-in-population fitness that fluctuates more often than in non-modulatory agents – the same is also true when comparing agents that evolve with random partners compared to coevolved partners.

Statistical Tests

After analysing the distributions and the statistical moments for each volatility metric, a Shapiro-Wilk test was conducted for each distribution to test for normality; as $p < 0.05$, each was found to be non-normal. Following the rationale for choosing statistical tests discussed in Section 6.5.1, one two-tailed ($m_n \neq m_m$) and two one-tailed ($m_n < m_m$, $m_n > m_m$) Wilcoxon Signed Rank statistical tests were conducted to ascertain whether behavioural

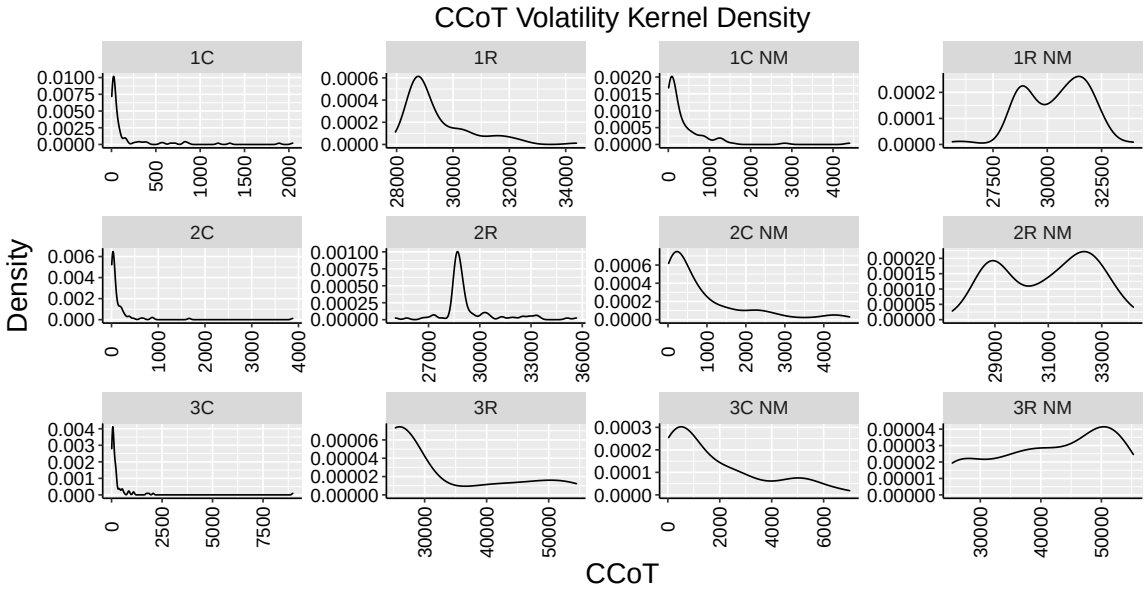


Figure 6.6: Kernel density estimations of the CCoT of agents that are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved (C) or random (R) partners. Experiments are repeated for non-modulatory and modulatory (NM) agents. *Note: the graph scales are not comparable due to the disparity in the densities between each distribution.*

plasticity has an impact on evolutionary volatility, by comparing non-modulatory (m_n) and modulatory (m_m) agents. The results are presented in Table 6.11.

For each of the median SDoT, CACoT and CCoT for each experiment, the modulatory and non-modulatory approaches were found to be significantly different ($p < 0.05$, $m_n \neq m_m$). An exception to this is when comparing the SDoT of agents that are evaluated on one environment with coevolved partners, as $p = 0.06919$; objectively this result is not significant, however the SDoT was found to be smaller in non-modulatory agents than their modulatory counterparts ($p = 0.03460$, $m_n < m_m$). Here, a 'small' effect size estimate r [58] indicates that behavioural plasticity has an observable – albeit weak – effect on the SDoT.

Non-modulatory agents were observed to have a lower SDoT and CACoT than modulatory agents when evolving with coevolved partners in each experiment, and a higher SDoT and CACoT when evolving with random partners; $p < 0.05$ in the one tailed tests with alternative hypotheses $m_n < m_m$ and $m_n > m_m$ respectively. The effect size estimate r corroborates these findings, as a negative r indicates the SDoT or CACoT is higher in modulatory agents than non-modulatory agents, whereas a positive r indicates the opposite. The exception to these findings is that the SDoT of non-modulatory agents that are evaluated on

Metric	Exp	Partner	Statistical Test Alternative Hypothesis			z	r
			$m_n \neq m_m$	$m_n < m_m$	$m_n > m_m$		
SDoT	1	Coevolved	6.919×10^{-2}	3.460×10^{-2} *	9.657×10^{-1}	−1.819	−0.1819 (S)
	1	Random	1.997×10^{-5} *	1	9.984×10^{-6} *	4.267	0.4267 (M)
	2	Coevolved	7.117×10^{-4} *	3.559×10^{-4} *	9.996×10^{-1}	−3.387	−0.3387 (M)
	2	Random	7.905×10^{-3} *	9.961×10^{-1}	3.952×10^{-3} *	2.658	0.2658 (S)
	3	Coevolved	7.482×10^{-4} *	3.741×10^{-4} *	9.996×10^{-1}	−3.373	−0.3373 (M)
	3	Random	3.250×10^{-5} *	1.625×10^{-5} *	1	−4.157	−0.4157 (M)
CACoT	1	Coevolved	1.371×10^{-6} *	6.856×10^{-7} *	1	−4.831	−0.4831 (M)
	1	Random	3.370×10^{-6} *	1	1.685×10^{-6} *	4.649	0.4649 (M)
	2	Coevolved	5.639×10^{-11}*	2.819×10^{-11}*	1	−6.555	−0.6555 (L)
	2	Random	1.292×10^{-5} *	1	6.458×10^{-6} *	4.363	0.4363 (M)
	3	Coevolved	4.072×10^{-13}*	2.036×10^{-13}*	1	−7.255	−0.7255 (L)
	3	Random	1.714×10^{-7} *	1	8.569×10^{-8} *	5.230	0.5230 (L)
CCoT	1	Coevolved	3.152×10^{-5} *	1.576×10^{-5} *	1	−4.164	−0.4164 (M)
	1	Random	3.130×10^{-5} *	1.565×10^{-5} *	1	−4.166	−0.4166 (M)
	2	Coevolved	5.493×10^{-11}*	2.746×10^{-11}*	1	−6.559	−0.6559 (L)
	2	Random	1.443×10^{-8} *	7.214×10^{-9} *	1	−5.670	−0.5670 (L)
	3	Coevolved	1.360×10^{-12}*	6.798×10^{-13}*	1	−7.090	−0.7090 (L)
	3	Random	2.423×10^{-8} *	1.212×10^{-8} *	1	−5.580	−0.5580 (L)

Table 6.11: Wilcoxon Signed Rank statistical tests comparing the volatility metrics of non-modulatory (m_n) and modulatory (m_m) agents that are evaluated on: (1) one, (2) two or (3) four environments in each experiment (Exp), with coevolved or random partners. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$). Effect sizes (r, to 4 S.F.) are presented with the z-score they are calculated from (Equation 3.3, $N = 100$), and are classed as small (S, $r \geq 0.1$), medium (M, $r \geq 0.3$), or large (L, $r \geq 0.5$) [58].

four environments is lower than in modulatory agents, when evolving with random partners ($p = 0.00001625$, $m_n < m_m$). This could be because fewer non-modulatory agents in these experiments are able to achieve their goal in at least one of the four of the environments they are evaluated on by the end of evolution, when compared with modulatory agents (Table 6.4); the SDoT would therefore be quite low compared to modulatory agents that achieve their goal more often, as fitness will fluctuate while goal-achieving behaviours are discovered, established and propagated throughout the population. Further, behavioural plasticity is shown to have a stronger effect on the CACoT of each experiment when agents evolve with random partners compared to coevolved partners due to a larger value of r, and tends to increase in line with the number of environments that agents evolving with either coevolved or random partners are evaluated on.

Non-modulatory agents are shown to have a lower CCoT than modulatory agents in all experiments ($p < 0.05$, $m_n \neq m_m$ and $m_n < m_m$); this observation is also supported by a negative effect size estimate r for each experiment. Further, behavioural plasticity generally has a stronger effect on the CCoT when agents evolve with coevolved partners compared

to random partners (the opposite is true in agents that evolve in one environment by a margin of 0.0002, which is negligible in terms of r); this indicates that the CCoT of agents that evolve with random partners is more similar than those that evolve with coevolved partners, and can be confirmed by looking at the median CCoT values in Table 6.10.

The findings in this section indicate that the type of partner an agent has (coevolved or random), as well as whether they are capable of behavioural plasticity, influence the amount of evolutionary volatility an agent experiences. Non-modulatory agents experience fewer fluctuations in fitness during evolution (CCoT) than modulatory agents, despite the type of partner they may evolve with. Non-modulatory agents also have a lower standard deviation in fitness over time, as well as a lower magnitude of fluctuations in fitness over time (SDoT and CACoT respectively) than modulatory agents when they evolve with a coevolved partner, however this reverses when the agents evolve with random partners. Interestingly, this shows that behavioural plasticity reduces the standard deviation in fitness during evolution, as well as the cumulative magnitude of changes in fitness when agents experience more variability (i.e. by evolving with random partners rather than coevolved partners). However, regardless of the type of partner an agent evolves with, non-modulatory agents experience fewer fluctuations in fitness during evolution; this could be a result of a lower percentage of non-modulatory agents being able to achieve their goal, and receiving lower fitnesses than modulatory agents.

Discussion and Summary

All three volatility metrics are useful to get a picture of how agents evolve, and what factors – such as behavioural plasticity, or the predictability of a partner's behaviour – affect the received fitness over time, to understand and predict how agents will behave in different environments or conditions. As the CCoT captures the number of times that an agent's fitness fluctuates over evolution, this is complemented by the CACoT, which captures the *magnitude* of those changes in fitness. The SDoT captures how much the fitness varies over time, and lower values in each of these metrics indicate less volatility. A low value in each volatility metric would indicate a low level of volatility, but this could also be influenced by a large number of agents that do not achieve their goal and thus experience little to no evolutionary volatility; it is therefore also useful to frame this volatility analysis in terms of the fitness received by agents as well.

There are two distinct conclusions about volatility to be made – one regarding the effect of behavioural plasticity on agents that learn in different environments, and the other regarding how the consistency of a partner affects the ability to achieve goals. Firstly, non-modulatory agents are found to have a lower CCoT than modulatory agents in each experiment (Table 6.11), but also a lower best-in-population fitness (Table 6.6). This indicates that behavioural plasticity increases the number of fluctuations in fitness during evolution compared to non-modulatory agents. Secondly, non-modulatory agents have a lower SDoT and CACoT than modulatory agents when evolving with coevolved partners, but generally a higher SDoT and CACoT with random partners. As non-modulatory agents have a higher SDoT and CACoT than modulatory agents, but lower CCoT when evolving with random partners, this would indicate that non-modulatory agents have a fitness that fluctuates less often but by larger magnitudes during evolution. Volatility therefore appears to also be affected by environmental factors such as whether the partner is coevolved (and thus has some element of predictability), or random. Additionally, evolutionary volatility, as captured by each of the three metrics used in this section, tends to increase as the number of environments that an agent is evaluated on at each generation is increased: agents that are evaluated on four environments experience the most volatility across each metric – both with and without neuromodulation – whereas agents that are evaluated on just one environment experience the least volatility.

These findings show that both behavioural plasticity and the predictability of the other agent in the environment can affect the volatility that agents experience during evolution; neuromodulation not only enables agents to achieve their goals more often, but can reduce evolutionary volatility when the behaviour or presence of another agent in the environment is unpredictable – even when the complexity of the task (through evaluating the agents on more environments at each generation) increases.

6.5.4 Analysing the Modulatory Neurons in the Neural Networks

As neuromodulation is seen to help agents to achieve their goals more often and receive a higher fitness, the question arises whether the configuration of modulatory neurons in the deliberative neural networks of the agents has an influence on goal-achievement. The deliberative neural network of the best-in-population agent after 500,000 generations of evolution is analysed for each of the 100 runs of all experiments. The most common configuration of

Experiment	Partner	Goal	L1	L2	L3	LT	Freq	Total
1	Coevolved	Yes	4	3	3	10	6	94
		No	-	-	-	-	-	6
	Random	Yes	4	3	3	10	5	86
		No	5	1	2	8	2	14
2	Coevolved	Yes	2	2	3	7	3	66
			3	2	1	6		
			3	3	2	8		
			4	2	2	8		
			4	3	2	9		
			5	2	2	9		
			5	3	3	11		
		No	2	3	2	7	2	34
			3	2	1	6		
			3	3	2	8		
			3	5	1	9		
			4	2	1	7		
			4	2	2	8		
			4	4	2	10		
	Random	Yes	3	3	2	8	4	67
		No	5	3	3	11	4	33
3	Coevolved	Yes	4	2	2	8	3	66
			4	4	3	11		
		No	3	2	1	6	2	34
			3	2	2	7		
			4	3	3	10		
	Random	Yes	3	3	2	8	5	86
		No	-	-	-	-	-	14

Table 6.12: The most common number of modulatory neurons evolved in each of the three layers of the deliberative neural networks (L1, L2, L3), and in total (LT), for agents evaluated on (1) one, (2) two or (3) four environments, that achieve their goal and those that do not, and that evolve with coevolved or random partners. The frequency that the configuration occurs is shown, as well as the total number of agents overall. A dash (-) indicates that no configuration occurred more than once, whilst multiple rows show configurations with the same frequency.

modulatory neurons in each of the neural network layers, and in total in the entire network, for each of the experiments is presented in Table 6.12; these are broken down into agents that achieve their goal in all environments they are evaluated on (either one, two or four), and those that do not.

These results show that no one configuration of modulatory neurons influences whether an agent is able to achieve its goal in the experiment or not. In fact, similarly to the findings presented in Section 5.5.4, the frequency of common neuron configurations is low compared to the total number of agents that do or do not achieve their goal. For some experiments, there are many configurations that occur in the population the same number of times; this shows that the same configuration can be evolved by agents that both achieve and do not

achieve their goal. For example, in agents that are evaluated on two environments and evolve with coevolved partners, two common configurations for all agents – regardless of whether they achieve their goal – for Layers 1, 2 and 3 respectively, are $\{3, 2, 1\}$ and $\{4, 2, 2\}$; these neurons are not necessarily arranged in the same order within the neural network. This supports the conclusion that the arrangement and number of modulatory neurons do not affect an agent's ability to achieve goals, no matter the type of environment that it may evolve in.

6.6 Discussion and Implications

In Chapter 5, the effect of behavioural plasticity on agent evolution is explored when agents evolve to solve a single- or multi-stage task; neuromodulation is observed to help evolution find higher-fitness solutions more often, with agents achieving their goals for a higher proportion of evolution than non-modulatory agents – but at the price of higher evolutionary volatility. In this chapter, the effect of behavioural plasticity on an agent's ability to achieve goals is further explored, by comparing how agents evolve to achieve their goals when experiencing a series of increasingly variable environments. Specifically, agents are evaluated on one, two or four environments at each generation, and may evolve with coevolved or random partners. In the natural world, behavioural plasticity can increase an individual's chance of survival in novel, variable environments, through enabling a genotype to vary its phenotype according to different environmental stimuli [196]. In this experimental study, behavioural plasticity is shown to improve the likelihood that an agent will evolve to achieve its goal, with modulatory agents evolving to achieve higher fitnesses – and more often – than their non-modulatory counterparts.

Evolving successful solutions is shown to be more difficult when agents experience higher levels of environmental variability; by evolving with a random partner at every generation, the environment becomes uncertain and behaviour successful in one context may be unsuccessful in another. However, whilst it is a challenge for evolution to find goal-achieving solutions when agents evolve with partners that are unpredictable, agents are observed to receive a higher fitness when they do in fact achieve their goal compared to when they learn in environments with more certainty.

A contributing factor to this is that agents that coevolve with other agents have a

tendency to evolve exploitative behaviour, whereas those that evolve with random partners have a tendency to evolve cooperative behaviour. Exploitative behaviour enables agents to receive the highest possible fitness as they do not endure any cost through bridge-building behaviour, however agents may then rely on the actions of others to achieve their goals. This can be seen clearly in Tables 6.2, 6.3 and 6.4, as more agents achieve their goals when coevolving with another agent by exploiting the actions of the other agent, than in a single-agent environment. Exploitative agents are unable to achieve any fitness when existing alone, meaning they have evolved to depend on the actions of others to achieve their goal.

In comparison, when agents are exposed to more experiences through enduring high levels of environmental variability, cooperative behaviour is more prevalent; a portion of agents that are able to achieve their goal alone go forward to cooperate with their partner agent, and thus receive a higher fitness than acting individually and incurring the entire cost of building the bridge on their own. Learning how to behave in an environment with a variety of random partners suggests that agents evolve the ability to consistently and reliably achieve a higher fitness through cooperation when situated in a shared environment, without that impacting their ability to achieve goals when alone. When observing human participants of an iterated Prisoner's Dilemma game, Fehl et al. [83] also found cooperation to be the favoured strategy when the anonymous partners in the game were prone to change – what Fehl et al. call a 'dynamic network'. In this chapter, the observation that cooperative behaviour can emerge in environments where agents have random partners – and also completely shift from the exploitative behaviour seen in environments with coevolved partners – is not only striking to see in Tables 6.2, 6.3 and 6.4, but deserving of more in-depth investigation in future work to explore the emergence of unintentional social dynamics.

As discussed in Section 2.5.3, Jolley and Channon [114] demonstrate that averaging the fitness that agents achieve when they are evaluated on many environments (in this case, ten) at once can improve overall performance. Experiencing a wide range of environmental scenarios can therefore be said to assist in the emergence of more generalised behaviour. The results presented in this chapter concur with this finding in the sense that evolving with random partners leads to a higher fitness being achieved than in agents that coevolve. However, even though the agents presented in this study are evaluated on an increasing number of environments in each experiment, this is not equivalent to the different numbers

of environmental evaluations seen in Jolley and Channon [114]; this is because the environments agents experience here differ to each other in terms of the perceived complexity (arising from being situated with another, or alone), which is not seen when agents are evaluated on multiple repetitions of the same environment.

The results of this experimental study and the consequent analyses show that evolving goal-achieving behaviour that is successful in a range of contexts is a difficult problem. While it appears to be harder to evolve successful solutions when partners are unpredictable, higher fitnesses become accessible through building resilience to changing environmental stimuli. Experiencing high levels of variability from changing environmental stimuli can therefore be more beneficial than evolving in a less uncertain environment, in terms of the fitness received and the behaviours evolved. Behavioural plasticity is seen to strengthen this effect further, as temporary, reversible changes in behaviour facilitate adaptations to novel environments and lead to higher fitness solutions being evolved. Interestingly, behavioural plasticity can actually reduce the amount of volatility experienced when evolving with random and therefore unpredictable partners; whilst the *frequency* of the fluctuations in fitness captured by the CCoT metric remain high, the *magnitude* of those changes instead reduces with behavioural plasticity. Therefore, even though fitness may change often when experiencing environmental uncertainty as a result of the unpredictable actions of random partners, behavioural plasticity is shown to have less of an effect on volatility as a whole.

6.7 Conclusion

In this chapter, the River Crossing Dilemma testbed is used to examine how evolving agents achieve the same task when environmental conditions or contexts vary to different extents, and how this environmental variability experienced can affect goal-achievement and evolution. Further, behavioural plasticity is one example of a beneficial quality in the natural world for surviving in novel environments [196]; as such, the effect that behavioural plasticity via activity-gating neuromodulation has on evolution and goal-achievement in dynamic environments is also explored.

The main findings and contributions of this chapter are:

- Environmental variability caused by the actions of others that constantly change is shown to positively affect agent evolution and goal-achievement; unintentional coop-

erative behaviour emerges in highly variable environments, shifting from exploitative behaviour observed in less variable environments. This relates to the first research question outlined in Chapter 1.

- Behavioural plasticity can enable agents to overcome variability, helping them to achieve their goals more often than non-plastic agents. This relates to the third research question outlined in Chapter 1.

- Evolutionary volatility is observed to decrease in plastic agents facing high levels of variability, in terms of the magnitude of changes in fitness over time, where variability is caused by the actions of random partners rather than coevolved partners. The three volatility metrics (SDoT, CACoT and CCoT) are used to measure this volatility, which relates to the second research question outlined in Chapter 1.

- There is a relationship between fitness and volatility for modulatory agents, where a trade-off exists in environments with low variability, but disappears when variability increases as volatility decreases. This relates to all three research questions outlined in Chapter 1.

When agents evolve with random partners, the context in which they are behaving, and attempting to achieve their goal in, continually changes from one generation to the next; consequently, it is harder for agents to evolve goal-achieving behaviour than when agents co-evolve with other agents. However, whilst evolving in highly variable environments is shown to be a challenge, agents are shown to receive a higher fitness than those in less variable environments as a result. Further, cooperative behaviour is found to be more prevalent in agents that evolve with random partners, whereas exploitative behaviour is more common in agents that coevolve with partners; this shows that variability over predictability is more useful for evolving goal-achieving behaviour in the face of variability or uncertainty. These findings are in line with those of Andras et al. [8], who demonstrate that environmental uncertainty can have a positive effect on promoting cooperative behaviour in agents. This shift in behaviour observed as a result of increasing the level of environmental variability that agents are exposed to is an especially interesting result; agents are incapable of perceiving or learning about others, or indeed their actions or the resulting consequences of those actions, so therefore any social dynamics are emergent and cannot be intentional.

Evolving in a more diverse environment is therefore shown to be more beneficial in terms of the fitnesses achieved after evolution, as exploitative agents may receive higher fitnesses when exploiting other agents, but cannot achieve their goals if they exist in an environment alone. By experiencing how the environment is affected by actions from a wide range of partners, agents are observed to receive higher fitnesses overall, and evolve more reliable behaviour that generalises across environments in which a partner may or may not exist. In fact, a relationship between variability and the benefit of plasticity is found, such that the benefit of plasticity is observed to get stronger as environmental variability increases – even when other agents in the environment are not known.

Extending on from this, behavioural plasticity is demonstrated to have a positive effect on agents in all areas of the study; agents capable of temporarily and reversibly changing their behaviour in response to changing environmental stimuli are more likely to achieve their goals, receive a higher fitness, and are successful for longer during evolution than those that are not. Furthermore, behavioural plasticity is shown to increase evolutionary volatility in all agents, in terms of the *number* of fluctuations in fitness over time, however the *magnitude* of those changes is actually shown to decrease when environments are highly variable (i.e. when agents evolve with a different partner at each generation). Therefore, activity-gating neuromodulation is shown to not only be beneficial in terms of agent fitness and goal-achievement, but the magnitude of evolutionary volatility experienced by agents reduces when environments are novel, uncertain, or variable. These results therefore provide evidence that behavioural plasticity via activity-gating neuromodulation is both a viable and useful mechanism that future systems could adopt in order to overcome this variability, and the negative effects of interference caused by other agents within shared environments.

Unforeseen environmental conditions and unpredictable interactions are characteristics of large, heterogeneous systems, with the frequency of unintended interactions increasing with system complexity and size. This study abstracts this problem to demonstrate that evolution in highly variable environments, caused by the unpredictable actions of others, can negatively affect a system's ability to achieve goals; goal-achieving behaviour must generalise across a wide range of environmental conditions and potential interactions with other systems that are unknown. Further, behavioural plasticity via activity-gating neuromodulation is shown to improve the prospect of goal-achievement in environments where the

predictability of the presence of other systems – and the actions in which they take – is constantly in flux. Finally, behavioural plasticity is observed to reduce evolutionary volatility in these highly variable environments, while increasing the expected fitness and the likelihood of goal-achievement.

This study demonstrates the importance of considering how systems may overcome unforeseen interactions or environmental conditions during the design process; modern technical systems are evermore situated in environments which are increasing in size and complexity, and thus variability. Whilst the results in this study highlight important issues to consider in system design, real-world technical systems comprise many interrelated components; a limitation of this study is that the abstracted agent-based model studies how only a maximum of two agents evolve and interact with one another. Future studies and extensions of this work will therefore explore the impact that many components within a system can have on evolution and goal-achievement in highly variable environments, and how the consequences of the resulting increase in the number of unintended interactions can be mitigated.

Chapter 7

Conclusion

This thesis presents three, comprehensive experimental studies that are designed to investigate the impact that interference can have on agents within a shared environment. As the computer systems of today are often composed of many colocated components, the River Crossing Dilemma testbed is introduced to explore issues of interference in a simplified model; the intention behind these studies is to gain an insight into the emergent phenomena that may arise in these complex, real-world systems, by studying how artificial agents evolve to achieve their goals in a simulated environment. Specifically, agents are evolved without a capacity to learn of other agents, meaning that they are unable to understand how the actions of others may affect themselves. This is important to study since it is infeasible for a system to obtain knowledge of all other actors within an environment, and the influence that those actors may have on the system itself [112]. Consequently, this is one novelty arising from this thesis, as other literature tends to focus on approaches to mitigate interference when agents are endowed with information about others. Since actors that are colocated are inherently socially situated and subject to interference from the actions of others, one of the objectives of the work presented within this thesis is to understand such interference at a fundamental level, as a prerequisite to the more complex social capabilities required for social awareness and intelligence [51, 28]. The results of the experimental studies show that acting in a more socially-oriented way can help agents to achieve their goals regardless of whether they exist in isolation, in a shared environment, or even when the presence of another agent is unpredictable. This therefore demonstrates that it is possible for agents to mitigate the consequences of interference and uncertainty arising from the

actions of other actors, *without* the explicit need for changing the way they operate when the context of the environment changes, for obtaining more information than is already available, or for modelling others within the environment in order to behave appropriately.

Three primary questions are asked in Chapter 1, intended to explore these issues of interference in detail. These surround: how agents can be affected by interference from unknown others; whether the effect of interference can be measured and compared; and finally, whether agents can mitigate the negative effects of interference despite being unable to reason about others and their actions. Since humans and other animals have evolved the ability to act socially in response to their shared environments, inspiration is taken from the fields of sociology, neuroscience and biology to investigate whether interference can be mitigated, and whether the goal-rationality seen in current systems is sufficient as these complex, social real-world systems grow larger. Specifically, taking actions that are socially-oriented and not necessarily goal-oriented is shown to be an effective way of reducing the evolutionary volatility experienced by agents, whilst not affecting the fitness received; plasticity as a result of neuromodulation is also shown to be beneficial in highly variable environments characterised by uncertainty and interference, improving both volatility and the fitness received by agents. These approaches demonstrate that acting in a socially-inspired, or socially-motivated manner can improve an agent's ability to achieve goals in shared environments – even without knowledge of others. The experimental studies therefore provide insight into the consequences that simple coexistence can have on the ability for systems to achieve goals, and examples of simple mechanisms that can be used to improve system performance when information about others may not be readily available.

7.1 Contributions Revisited

To summarise, the major contributions of this thesis are as follows:

- The River Crossing Dilemma testbed, designed to explore social phenomena, the consequences of interference, and how agents achieve goals in environments of arbitrary complexity. Specifically, this testbed is used to explore how agents may evolve in shared environments without any knowledge, or ability to acquire such knowledge, of other agents and the effect that their actions may have on themselves. This testbed is thus used to explore the first and third research questions detailed in Chapter 1.

- Three volatility metrics – SDoT, CACoT, and CCoT – that can be used to quantify the frequency and/or magnitude of changes in fitness during agent evolution, such that the effects of interference can be measured and compared between agents in different environments, or of different designs; these metrics can be used to explore the second research question in Chapter 1.

- A study that concludes goal-rationality is sufficient for achieving goals in shared environments when experiencing interference; however, the traditional social action proposed can achieve similar results whilst reducing evolutionary volatility. This demonstrates the viability of more socially-oriented behaviour when systems or agents exist in shared environments, and highlights that a trade-off can be made between the ability to achieve goals and the volatility experienced. This study therefore explores all three research questions presented in Chapter 1.

- Two studies which demonstrate that behavioural plasticity – the temporary, reversible changes in behaviour achieved in this case with a novel approach to neuromodulation – enables agents to achieve their goals more often when facing higher levels of interference and uncertainty. Plasticity is also shown to reduce the evolutionary volatility experienced by agents in highly variable environments, whilst improving the ability to achieve goals compared to non-plastic agents. Further, unintentional cooperation is shown to emerge in highly variable environments which is more prevalent in plastic agents, whereas self-interested behaviour is more common in less variable environments. These studies explore all three research questions detailed in Chapter 1.

- Three experimental studies that explore the consequences of interference and how they can be mitigated by observing: how agents evolve to achieve their goals in isolation, compared to when they exist with one other; how evolving in an environment that can change from solitary to shared can affect evolution and goal-achievement; how the task complexity can affect evolution and the interference experienced in these environments; and how the predictability of a partner agent's actions, and the resulting level of uncertainty or variability within the environment, can affect evolution and goal-achievement. Overall, each study is designed to explore each of the three research questions presented in Chapter 1, with each study focusing on investigating the effects that interference can have on agents in different social situations.

The three experimental studies presented in Chapters 4, 5 and 6 have been designed to answer the three research questions detailed in Chapter 1. The first research question is formulated around understanding how agents in shared environments are affected by interference from the actions of others. The results obtained from these studies demonstrate that interference can either be beneficial to agents by helping them to achieve a higher fitness through cooperating with or exploiting the other, or it can cause goal-achieving behaviour to become unreliable as the actions of the other agent changes the environment in ways that the agent cannot understand. The second research question asks how the effect of this observed interference can be quantified; as such, three volatility metrics are introduced, utilised in each study to measure the effect that interference can have on the evolution of agents in each experiment. These metrics are then used to answer the third research question, which concerns how agents can be designed to mitigate the effects of interference – which can be quantified and compared using these three metrics.

The studies conducted to investigate the research questions in this thesis demonstrate that the effect of interference can be great and catastrophic in terms of developing or sustaining the ability to achieve individual goals, but also that simple methods can be employed to help agents mitigate the effect of interference *without* having to learn about the existence of others. This is important because other work that explores interference focuses on how to learn about others or the existence of interference itself, or alternatively on mitigating interference when others are already known. Acquiring knowledge about others in the environment is not always possible, either immediately or at all; this depends on the agent's ability to learn about or model others, or the processing power required to obtain such information depending on the size of the system [112]. Consequently, this thesis investigates the often overlooked scenario of how agents are affected by interference *without* such knowledge of others, by observing agent behaviour in environments inhabited by either one or two agents. This again is important because in the real-world, complex sociotechnical systems with many parts exist in inherently social environments, where interference from unintended interactions increases with system size and the number of components [101]; these systems should be capable of maintaining goal-achieving behaviour even in the face of uncertainty, and in situations where others causing interference are not yet known or modelled adequately enough to make appropriate decisions.

As highlighted by Bellman et al. [28], social awareness and the ability to make decisions

informed by social information are required for future sociotechnical systems to act appropriately within their environments; however, these systems should still be able to perform consistently whilst learning about these social concepts and the way that they can be influenced by others – or indeed when the existence of other systems is not yet known, or knowable. This thesis shows that simple methods inspired by sociology, neuroscience and biology can be used to realise this endeavour, by enabling systems to mitigate unknown interference as a step towards the design of socially intelligent, and socially-sensitive machines. The studies conducted within this thesis, designed to investigate these problems in social environments, therefore contribute to the wider understanding of the consequences that interference can have on agents that simply pursue individual goals within a shared environment. In doing this, the results of these experiments demonstrate the importance of considering the implications of interference when designing systems that may operate in shared and uncertain environments; this is because goal-achievement can be hindered if neither interference nor the actions of others are understood, which can have potentially catastrophic consequences in the real-world. As real-world systems become larger, they will be more prone to interacting with many other artificial or human actors in the environment – intentionally or not. These other human or machine actors may enter, leave or behave in the environment unpredictably; if these systems can be endowed with the ability to mitigate the consequences of interference without explicitly requiring any additional modelling, processing, or information, we may find ourselves closer to realising socially intelligent systems that are capable of acting appropriately in real-world environments that are characterised by uncertainty and heterogeneity.

7.2 Limitations, Direction and Future Work

As the real-world sociotechnical systems that this thesis is motivated by are growing evermore complex, it is unrealistic that these systems or the components they comprise will exist in isolation – or even colocated with only one other actor, for example. Although the study of how at most one or two agents pursue individual goals in a shared environment is important, and has given valuable insight into the effects of interference in the minimal sense, observing only one or two agents is also a limitation of the experimental studies conducted within this thesis. The number of interactions, and thus the severity of the

consequences of interference, will increase in line with the number of components within a system [101]. As highlighted by Heylighen [107], the actions of individuals can often cause a ripple effect, meaning that the consequences of interference may initially be felt in local regions of the environment, but may escalate and cause chain reactions that affect actors on a global scale. As a result, one future research direction is to investigate the effects of interference in environments inhabited by many actors (in this case, more than two); this is because the consequences of the actions of one agent may be magnified in densely populated environments, and this cannot be understood fully by observing only two actors.

The three metrics used for analysing the evolutionary volatility that agents experience are shown to be useful for gathering an insight into how resilient agents are to varying environmental stimuli, resulting from the actions of others. However, a limitation of these metrics is that they capture *all* changes in fitness experienced by agents during evolution; they don't distinguish whether these changes are positive or negative, which would indicate a benefit or detriment to goal-achieving behaviour respectively. An area for future investigation would therefore be to also analyse how often an agent loses or gains the ability to achieve its goal during evolution. This would be important as the current metrics capture evolutionary changes be those positive or negative, however what might be more insightful is to also capture the *implications* of these changes. Further, the three metrics must be analysed in the context of the fitness received by agents in order to get a clear picture of how agents evolve in these isolated or shared environments. A low value in each volatility metric would for example indicate resilience and that the fitness does not change often during evolution, but could also indicate that agents simply consistently do not achieve their goals. Additional metrics that measure when the goal-achieving fitness threshold is passed (either for benefit or detriment) could alleviate some of this ambiguity, whilst also providing a deeper insight into the behaviour of these agents when experiencing interference.

Modern-day sociotechnical systems are characterised by heterogeneous components, which may potentially have different goals, abilities, and knowledge. An interesting future avenue of research would be to explore how a mixture of different types of actors within a shared environment would interact and affect one another. This would strengthen the need for social awareness [28], since some actors may possess the ability to model and reason about others, whereas some may not; regardless, each component within such a system would need the ability to act appropriately given their inherently social circumstances, to

avoid catastrophes [46]. By studying this in a simplified agent-based model, the dynamics that emerge in such rich and diverse environments can begin to be understood, which could better inform the designers of these complex systems in the real-world.

A longer-term trajectory for this line of research would be to explore how artificial agents or systems would interact in a shared environment with humans. Similarly to the experiments discussed in this thesis, a simplified model could be used to explore the interactions between human and machine, whilst also exploring how interference can be mitigated between these actors on a larger scale compared to the work presented here. The primary focus of this line of research, as with this thesis, would be on how actors within this model are able to mitigate interference without necessarily knowing about others, the effects that the actions of others can have, or the abilities of others – as a step towards social intelligence. As this model would theoretically be more complex than the one used within this thesis since there would be more than two actors (complexity can increase as a result of the interactions and interference between actors [191]), more diverse and complex questions could be explored than those presented in this thesis, for example: how might agents (or even physical entities such as robots, rather than simulated agents) develop the ability to perceive others that they have not encountered before? How may these actors mitigate interference that originates from a variety of different sources, be that natural or artificial? How might they do this without necessarily possessing knowledge of others, or the existence of the interference itself? How may the actions of human actors within the system unintentionally affect the way that artificial components (or even other humans) operate or achieve their goals? Is it possible for actors to model the capabilities of others, such that this knowledge can inform future decisions about how to act most appropriately given the social circumstances in which they exist? Further, would it be possible for an actor to adapt or change their behaviour based on the capabilities of others, for example in the extreme cases where another actor may not be able to model others at all, or may be better equipped for the task at hand? This final question could influence the emergence of social dynamics such as cooperation, competition, or coordination in order to achieve collective goals. The studies conducted within this thesis explore the consequences of interference, and how this manifests when agents share an environment when they have no knowledge of others; this provides a foundation on which the questions above can be explored, which are all viable avenues for future research into designing artificial systems that can operate in

social environments – and ultimately how socially intelligent systems may be realised. This thesis demonstrates the importance of studying the effects of interference at a fundamental level *before* these complex questions can be asked, contributing valuable insights into the implications that simply existing in a shared space can have for the goal-achievement and evolution of artificial agents. If systems can be designed to mitigate interference without requiring such information or knowledge about others, this could lead to the realisation of socially intelligent systems capable of operating consistently in environments shared by a variety of both human and artificial actors – enabling them to not only integrate into their local environment, but into the fabric of society as well.

Appendix A

Statistical Analysis Setup

Statistical analyses have been conducted in the experimental studies using the R programming language [173], and RStudio. Below is a list of the packages and functions used.

R packages used for statistical analyses and visualisations of data:

- `ggplot2` [229]: This package is used for creating graphics and visualising data, by mapping variables to aesthetics.
- `moments` [123]: This package provides functions to calculate statistical moments of a dataset (e.g. mean, variance, skewness and kurtosis).
- `rcompanion` [142]: This package provides core analysis functionality and datasets, including the `wilcoxonZ` function described below.
- `reshape2` [228]: This package is used to transform data into more suitable formats for analysis and visualisation.
- `tidyverse` [230]: This package provides a variety of functionality, including reshaping of data, as well as visualisation (`ggplot2`).
- `viridis` [95]: This package provides different colour palettes, and aims to improve readability of visualisations for people with various forms of colour vision deficiency and thus making data more accessible.

R functions used for statistical analysis:

- `wilcox.test(x, y, paired = "true")`: A two-tailed, paired Wilcoxon Signed Rank statistical test can be used to ascertain whether there is any significant difference

between the medians of two distributions. One-tailed tests can also be conducted by including the parameter `alternative = "less"` or `alternative = "greater"`, to test the alternative hypotheses $x < y$ and $x > y$ respectively.

- `shapiro.test(x)`: The Shapiro-Wilk test is used to test whether a distribution is normal or non-normal.

- `wilcoxonZ(x, y, paired = "true")`: The z-statistic for a Wilcoxon paired test can be used to calculate the effect size r, or relationship, between two variables ($r = z/\sqrt{N}$, where N is the number of samples).

- `mean(x), median(x), var(x), skewness(x), kurtosis(x)`: The mean, median, variance, skewness and kurtosis can be calculated to describe the shape of a distribution.

- `summary(x)`: Summary statistics for a distribution, such as the minimum, maximum, mean, and first, second and third quartiles can indicate the dispersion of values in the distribution.

Appendix B

Shapiro-Wilk Statistical Tests

Shapiro-Wilk tests indicate whether a distribution is normal or non-normal. The p-values for each set of tests in each chapter are detailed below.

B.1 Normality Tests: Traditional Action and Evolutionary Volatility

Below are the results for the Shapiro-Wilk tests conducted in the study presented in Chapter 4. Table B.1 presents the results of the statistical tests conducted on the SDoT, CACoT and CCoT distributions for agents using goal-rational, traditional and random action, evolving either alone or together. Table B.2 then presents the results for the fitness distributions of the same agents, whilst Table B.3 presents the results for the distributions of the number of generations that agents successfully achieve their goals during evolution.

Metric	Experiment	Action Type	p
SDoT	Alone	G	2.200×10^{-16}*
		GT	2.200×10^{-16}*
		GR	4.963×10^{-1}
	Together	G	7.961×10^{-12}*
		GT	7.532×10^{-14}*
		GR	9.603×10^{-2}
CACoT	Alone	G	2.200×10^{-16}*
		GT	2.200×10^{-16}*
		GR	4.607×10^{-8} *
	Together	G	2.200×10^{-16}*
		GT	2.200×10^{-16}*
		GR	1.069×10^{-10}*
CCoT	Alone	G	2.200×10^{-16}*
		GT	2.200×10^{-16}*
		GR	4.607×10^{-8} *
	Together	G	2.200×10^{-16}*
		GT	2.200×10^{-16}*
		GR	2.226×10^{-11}*

Table B.1: Shapiro-Wilk statistical tests for normality, for the volatility of agents that evolve with goal-rational, traditional, or random action, when alone or together. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$), indicating non-normality.

Experiment	Action Type	p
Alone	G	2.200×10^{-16}*
	GT	2.200×10^{-16}*
	GR	-
Together	G	7.842×10^{-12}*
	GT	1.007×10^{-12}*
	GR	4.705×10^{-14}*

Table B.2: Shapiro-Wilk statistical tests for normality, for the fitness of agents that evolve with goal-rational, traditional, or random action, when alone or together. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$), indicating non-normality.

Experiment	Action Type	p
Alone	G	6.618×10^{-16}*
	GT	2.220×10^{-16}*
	GR	1.132×10^{-9} *
Together	G	1.356×10^{-11}*
	GT	3.284×10^{-12}*
	GR	1.128×10^{-10}*

Table B.3: Shapiro-Wilk statistical tests for normality, for the number of generations that agents achieve their goals during evolution when using goal-rational, traditional, or random action, when alone or together. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$), indicating non-normality.

B.2 Normality Tests: Neuromodulation and Evolutionary Volatility

Below are the results for the Shapiro-Wilk tests conducted in the study presented in Chapter 5. Table B.4 presents the results of the statistical tests conducted on the fitness distributions for modulatory and non-modulatory agents that evolve to solve a single- or multi-stage task, either alone or together with another agent. Table B.5 then presents the results for the distributions of the number of generations that agents successfully achieve their goals during evolution. Finally, Table B.3 presents the results for the volatility distributions of agents.

Experiment	Task	Neuromodulation	p
Alone	Single-Stage	No	$1.133\times10^{-14*}$
		Yes	$2.200\times10^{-16*}$
	Multi-Stage	No	$7.104\times10^{-15*}$
		Yes	$2.200\times10^{-16*}$
Together	Single-Stage	No	$1.044\times10^{-10*}$
		Yes	$9.758\times10^{-13*}$
	Multi-Stage	No	$9.848\times10^{-11*}$
		Yes	$2.335\times10^{-12*}$
Continued Evolution	Single-Stage	No	$9.179\times10^{-10*}$
		Yes	$3.807\times10^{-13*}$
	Multi-Stage	No	$2.149\times10^{-10*}$
		Yes	$1.598\times10^{-13*}$

Table B.4: Shapiro-Wilk statistical tests for normality, for the fitness of non-modulatory or modulatory agents that evolve to solve a single- or multi-stage task. Agents evolve alone, together or with continued evolution. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$), indicating non-normality.

Experiment	Task	Neuromodulation	p
Alone	Single-Stage	No	$5.960\times10^{-13*}$
		Yes	$2.496\times10^{-10*}$
	Multi-Stage	No	$4.333\times10^{-14*}$
		Yes	$1.255\times10^{-10*}$
Together	Single-Stage	No	$1.478\times10^{-11*}$
		Yes	$4.183\times10^{-14*}$
	Multi-Stage	No	$1.875\times10^{-9\ *}$
		Yes	$1.019\times10^{-10*}$
Continued Evolution	Single-Stage	No	$5.305\times10^{-14*}$
		Yes	$2.200\times10^{-16*}$
	Multi-Stage	No	$5.903\times10^{-14*}$
		Yes	$2.200\times10^{-16*}$

Table B.5: Shapiro-Wilk statistical tests for normality, for the number of generations that non-modulatory or modulatory agents achieve their goals when evolving to solve a single- or multi-stage task. Agents evolve alone, together or with continued evolution. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$), indicating non-normality.

Metric	Experiment	Task	Neuromodulation	p
SDoT	Alone	Single-Stage	No	8.540×10^{-14}*
			Yes	1.363×10^{-7} *
		Multi-Stage	No	9.834×10^{-15}*
			Yes	1.512×10^{-8} *
	Together	Single-Stage	No	2.565×10^{-12}*
			Yes	1.441×10^{-7} *
		Multi-Stage	No	4.913×10^{-10}*
			Yes	9.884×10^{-7} *
	Continued Evolution	Single-Stage	No	8.748×10^{-9} *
			Yes	1.841×10^{-5} *
		Multi-Stage	No	1.538×10^{-10}*
			Yes	1.595×10^{-6} *
CACoT	Alone	Single-Stage	No	2.200×10^{-16}*
			Yes	2.200×10^{-16}*
		Multi-Stage	No	1.106×10^{-15}*
			Yes	4.854×10^{-14}*
	Together	Single-Stage	No	4.357×10^{-10}*
			Yes	1.561×10^{-12}*
		Multi-Stage	No	2.200×10^{-16}*
			Yes	5.105×10^{-14}*
	Continued Evolution	Single-Stage	No	2.196×10^{-15}*
			Yes	5.153×10^{-14}*
		Multi-Stage	No	2.200×10^{-16}*
			Yes	3.338×10^{-14}*
CCoT	Alone	Single-Stage	No	2.200×10^{-16}*
			Yes	2.200×10^{-16}*
		Multi-Stage	No	1.101×10^{-15}*
			Yes	4.962×10^{-14}*
	Together	Single-Stage	No	6.204×10^{-13}*
			Yes	1.135×10^{-13}*
		Multi-Stage	No	2.200×10^{-16}*
			Yes	1.608×10^{-15}*
	Continued Evolution	Single-Stage	No	6.923×10^{-16}*
			Yes	2.029×10^{-13}*
		Multi-Stage	No	2.200×10^{-16}*
			Yes	4.883×10^{-14}*

Table B.6: Shapiro-Wilk statistical tests for normality, for the volatility of non-modulatory or modulatory agents that evolve to solve a single- or multi-stage task. Agents evolve alone, together or with continued evolution. p-values (to 4 S.F.) are marked with an asterisk (*) if significant ($p < 0.05$), indicating non-normality.

B.3 Normality Tests: Environmental Variability and Evolutionary Volatility

Below are the results for the Shapiro-Wilk tests conducted in the study presented in Chapter 6. Table B.7 presents the results of the statistical tests conducted on the fitness distributions for modulatory and non-modulatory agents that are evaluated on one, two, or four environments at each generation, with coevolved or random partners. Table B.8 then presents the results for the distributions of the number of generations that agents successfully achieve their goals during evolution, and Table B.9 presents the results for the volatility distributions of agents.

Experiment	Partner	NM	p
1	Coevolved	N	$9.848\times10^{-11*}$
		Y	$2.335\times10^{-12*}$
	Random	N	$2.397\times10^{-12*}$
		Y	$3.968\times10^{-14*}$
2	Coevolved	N	$1.607\times10^{-12*}$
		Y	$5.822\times10^{-16*}$
	Random	N	$1.650\times10^{-11*}$
		Y	$1.476\times10^{-14*}$
3	Coevolved	N	$8.318\times10^{-9\ *}$
		Y	$1.441\times10^{-13*}$
	Random	N	$5.732\times10^{-13*}$
		Y	$6.035\times10^{-13*}$

Table B.7: Shapiro-Wilk tests for normality for the fitnesses of agents after evolving for 500,000 generations; agents are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved or random partners. p-values significant at $p < 0.05$, indicating non-normality, are marked with an asterisk (*).

Experiment	Partner	NM	p
1	Coevolved	N	1.875×10^{-9} *
		Y	1.019×10^{-10}*
	Random	N	5.749×10^{-13}*
		Y	2.024×10^{-7} *
2	Coevolved	N	7.831×10^{-14}*
		Y	7.093×10^{-11}*
	Random	N	2.484×10^{-14}*
		Y	3.984×10^{-10}*
3	Coevolved	N	1.242×10^{-15}*
		Y	2.494×10^{-10}*
	Random	N	2.385×10^{-14}*
		Y	6.729×10^{-7} *

Table B.8: Shapiro-Wilk tests for normality for the number of generations that agents achieve their goals during 500,000 generations of evolution; agents are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved or random partners. p-values significant at $p < 0.05$, indicating non-normality, are marked with an asterisk (*).

Metric	Experiment	Partner Type	Neuromodulation	p
SDoT	1	Coevolved	No	4.913×10^{-10}*
			Yes	4.913×10^{-10}*
		Random	No	2.880×10^{-14}*
			Yes	2.594×10^{-6} *
	2	Coevolved	No	2.102×10^{-12}*
			Yes	1.545×10^{-5} *
		Random	No	6.501×10^{-15}*
			Yes	6.729×10^{-8} *
	3	Coevolved	No	5.077×10^{-10}*
			Yes	4.814×10^{-4} *
		Random	No	1.535×10^{-10}*
			Yes	4.929×10^{-6} *
CACoT	1	Coevolved	No	2.220×10^{-16}*
			Yes	5.105×10^{-14}*
		Random	No	5.187×10^{-13}*
			Yes	4.214×10^{-7} *
	2	Coevolved	No	2.220×10^{-16}*
			Yes	2.418×10^{-10}*
		Random	No	3.467×10^{-13}*
			Yes	2.679×10^{-9} *
	3	Coevolved	No	2.220×10^{-16}
			Yes	1.800×10^{-9}
		Random	No	2.380×10^{-13}
			Yes	5.899×10^{-6} *
CCoT	1	Coevolved	No	2.220×10^{-16}*
			Yes	1.608×10^{-15}*
		Random	No	3.977×10^{-10}*
			Yes	7.555×10^{-4} *
	2	Coevolved	No	2.220×10^{-16}*
			Yes	1.133×10^{-11}*
		Random	No	1.033×10^{-10}*
			Yes	3.606×10^{-5} *
	3	Coevolved	No	2.220×10^{-16}*
			Yes	4.077×10^{-9} *
		Random	No	2.162×10^{-13}*
			Yes	1.111×10^{-5} *

Table B.9: Shapiro-Wilk tests for normality for the volatility of agents after evolving for 500,000 generations; agents are evaluated on (1) one, (2) two or (3) four environments in each experiment, with either coevolved or random partners. p-values significant at $p < 0.05$, indicating non-normality, are marked with an asterisk (*).

List of References

[1] L. F. Abbott. Modulation of Function and Gated Learning in a Network Memory. *Proceedings of the National Academy of Sciences of the United States of America*, 87(23):9241–9245, 1990. ISSN: 00278424. doi: https://doi.org/10.1073/pnas.87.23.9241.

[2] L. F. Abbott and S. B. Nelson. Synaptic Plasticity: Taming the Beast. *Nature Neuroscience*, 3(11): 1178–1183, 2000.

[3] R. Adolphs. Social Cognition and the Human Brain. *Trends in Cognitive Sciences*, 3(12):469–479, 12 1999. ISSN: 1879-307X. doi: https://doi.org/10.1016/S1364-6613(99)01399-6.

[4] W. Aguilar, G. Santamaría-Bonfil, T. Froese, and C. Gershenson. The Past, Present, and Future of Artificial Life. *Frontiers in Robotics and AI*, 1, 2014. ISSN: 2296-9144. doi: https://doi.org/10.3389/frobt.2014.00008.

[5] M. N. Aldelaimi, M. A. Hossain, and M. F. Alhamid. Building Dynamic Communities of Interest for Internet of Things in Smart Cities. *Sensors*, 20(10), 2020. ISSN: 1424-8220. doi: https://doi.org/10.3390/s20102986.

[6] W. L. Alexander Pokahr, Lars Braubach. Jadex: Implementing a BDI-Infrastructure for JADE Agents. *EXP – In Search of Innovation (Special Issue on JADE)*, 3(3):76–85, 2003.

[7] J. M. Anderies, M. A. Janssen, A. Lee, and H. Wasserman. Environmental Variability and Collective Action: Experimental Insights from an Irrigation Game. *Ecological Economics*, 93:166–176, 2013. ISSN: 0921-8009. doi: https://doi.org/10.1016/j.ecolecon.2013.04.010.

[8] P. Andras, J. Lazarus, and G. Roberts. Environmental Adversity and Uncertainty Favour Cooperation. *BMC Evolutionary Biology*, 7(1):240–247, 2007.

[9] D. E. Asher, A. Zaldivar, B. Barton, A. A. Brewer, and J. L. Krichmar. Reciprocity and Retaliation in Social Games With Adaptive Agents. *IEEE Transactions on Autonomous Mental Development*, 4 (3):226–238, 2012. doi: https://doi.org/10.1109/TAMD.2012.2202658.

[10] R. Axelrod and W. D. Hamilton. The Evolution of Cooperation. *Science (New York, N.Y.)*, 211 (4489):1390–1396, 3 1981. doi: https://doi.org/10.1126/science.7466396.

[11] F. A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. J. Filho, R. Lent, and S. Herculano-Houzel. Equal Numbers of Neuronal and Nonneuronal Cells Make the Human Brain an Isometrically Scaled-Up Primate Brain. *Comparative Neurology*, 513(5):532–541, 2009. doi: https://doi.org/10.1002/cne.21974.

[12] T. Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, 1996.

[13] M. Bakhshipour, M. Jabbari Ghadi, and F. Namdari. Swarm Robotics Search & Rescue: A Novel Artificial Intelligence-Inspired Optimization Approach. *Applied Soft Computing*, 57:708–726, 8 2017. ISSN: 1568-4946. doi: https://doi.org/10.1016/J.ASOC.2017.02.028.

[14] C. M. Barnes, K. Bellman, J. Botev, A. Diaconescu, L. Esterle, C. Gruhl, C. Landauer, P. R. Lewis, P. R. Nelson, A. Stein, C. Stewart, and S. Tomforde. CHARIOT – Towards a Continuous High-Level Adaptive Runtime Integration Testbed. In *Proceedings of the IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 52–55. IEEE, 2019. doi: https://doi.org/10.1109/FAS-W.2019.00026.

[15] C. M. Barnes, A. Ekárt, and P. R. Lewis. Social Action in Socially Situated Agents. In *Proceedings of the IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 97–106. IEEE, 2019. doi: https://doi.org/10.1109/SASO.2019.00021.

[16] C. M. Barnes, L. Esterle, and J. N. A. Brown. "When You Believe in Things That You Don't Understand": The Effect of Cross-Generational Habits on Self-Improving System Integration. In *Proceedings of the IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 28–31. IEEE, 2019. doi: https://doi.org/10.1109/FAS-W.2019.00020.

[17] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Coevolutionary Learning of Neuromodulated Controllers for Multi-Stage and Gamified Tasks. In *Proceedings of the IEEE 1st International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 129–138. IEEE, 2020. doi: https://doi.org/10.1109/ACSOS49614.2020.00034.

[18] C. M. Barnes, A. Ekárt, and P. R. Lewis. Beyond Goal-Rationality: Traditional Action Can Reduce Volatility in Socially Situated Agents. *Future Generation Computer Systems*, 113:579–596, 2020. doi: https://doi.org/10.1016/j.future.2020.07.033.

[19] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Behavioural Plasticity Can Help Evolving Agents in Dynamic Environments But at the Cost of Volatility. *ACM Transactions on Autonomous Adaptive Systems*, 2021. In Press.

[20] C. M. Barnes, A. Ekárt, K. O. Ellefsen, K. Glette, P. R. Lewis, and J. Tørresen. Evolving Neuromodulated Controllers in Variable Environments. In *Proceedings of the IEEE 2nd International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, 2021. In Press.

[21] C. M. Barnes, A. Ghouri, and P. R. Lewis. Explaining Evolutionary Agent-Based Models via Principled Simplification. *Artificial Life*, 27(3), 2021. In Press.

[22] L. Barrett, P. Henzi, and D. Rendall. Social Brains, Simple Minds: Does Social Complexity Really Require Cognitive Complexity? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):561–575, 4 2007. doi: https://doi.org/10.1098/rstb.2006.1995.

[23] S. Beaulieu, L. Frati, T. Miconi, J. Lehman, K. O. Stanley, J. Clune, and N. Cheney. Learning to Continually Learn. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI)*, pages 992–1001. IOS Press, 2020. doi: https://doi.org/10.3233/FAIA200193.

[24] G. S. Becker. Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology. *Journal of Economic Literature*, 14(3):817–826, 1976, URL: https://www.jstor.org/stable/2722629.

[25] M. A. Bedau. Artificial Life. In M. Matthen and C. Stephens, editors, *Philosophy of Biology*, Handbook of the Philosophy of Science, pages 585–603. North-Holland, Amsterdam, 2007. doi: https://doi.org/10.1016/B978-044451543-8/50027-7.

[26] F. Bellifemine, A. Poggi, and G. Rimassa. JADE – A FIPA-Compliant Agent Framework. In *Proceedings of PAAM*, volume 99, pages 97–108. London, 1999.

[27] K. Bellman, S. Tomforde, and R. P. Würtz. Interwoven Systems: Self-Improving Systems Integration. In *Proceedings of the IEEE 8th International Conference on Self-Adaptive and Self-Organizing Systems Workshops*, pages 123–127. IEEE, 2014. doi: https://doi.org/10.1109/SASOW.2014.21.

[28] K. Bellman, J. Botev, H. Hildmann, P. R. Lewis, S. Marsh, J. Pitt, I. Scholtes, and S. Tomforde. Socially-Sensitive Systems Design: Exploring Social Potential. *IEEE Technology and Society Magazine*, 36(3):72–80, 2017. ISSN: 02780097. doi: https://doi.org/10.1109/MTS.2017.2728727.

[29] K. Bellman, J. Botev, A. Diaconescu, L. Esterle, C. Gruhl, C. Landauer, P. R. Lewis, A. Stein, S. Tomforde, and R. P. Würtz. Self-Improving System Integration – Status and Challenges After Five Years of SISSY. In *Proceedings of the IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 160–167. IEEE, 2018. doi: https://doi.org/10.1109/FAS-W.2018.00042.

[30] K. L. Bellman, C. Gruhl, C. Landauer, and S. Tomforde. Self-Improving System Integration – On a Definition and Characteristics of the Challenge. In *Proceedings of the IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 1–3. IEEE, 2019. doi: https://doi.org/10.1109/FAS-W.2019.00014.

[31] E. Z. Berglund. Using Agent-Based Modeling for Water Resources Planning and Management. *Journal of Water Resources Planning and Management*, 141(11):04015025, 2015. doi: https://doi.org/10.1061/(ASCE)WR.1943-5452.0000544.

[32] J. J. Bird, C. M. Barnes, C. Premebida, A. Ekárt, and D. R. Faria. Country-Level Pandemic Risk and Preparedness Classification Based on COVID-19 Data: A Machine Learning Approach. *PLOS ONE*, 15(10):1–20, 10 2020. doi: https://doi.org/10.1371/journal.pone.0241332.

[33] J. J. Bird, C. M. Barnes, L. J. Manso, A. Ekárt, and D. R. Faria. Fruit Quality and Defect Image Classification with Conditional GAN Data Augmentation. *Scientia Horticulturae*, 2021. In Press.

[34] E. Bonabeau. Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287, 2002. ISSN: 0027-8424. doi: https://doi.org/10.1073/pnas.082080899.

[35] J. M. Borg and A. Channon. Evolutionary Adaptation to Social Information Use Without Learning. In G. Squillero and K. Sim, editors, *Applications of Evolutionary Computation*, pages 837–852. Springer, 2017. ISBN: 978-3-319-55849-3. doi: https://doi.org/10.1007/978-3-319-55849-3_54.

[36] J. M. Borg and A. Channon. The Effect of Social Information Use Without Learning on the Evolution of Social Behavior. *Artificial Life*, 26(4):431–454, 02 2021. ISSN: 1064-5462. doi: https://doi.org/10.1162/artl_a_00328.

[37] J. M. Borg, A. Channon, C. Day, et al. Discovering and Maintaining Behaviours Inaccessible to Incremental Genetic Evolution Through Transcription Errors and Cultural Transmission. In *Advances in Artificial Life: Proceedings of the 11th European Conference on the Synthesis and Simulation of Living Systems (ECAL 2011)*, pages 101–108. MIT Press, 2011. doi: https://doi.org/10.7551/978-978-0-262-29714-1-ch019.

[38] T. Bouron, J. Ferber, and F. Samuel. MAGES: A Multiagent Testbed for Heterogeneous Agents. In *Decentralised Artificial Intelligence*, volume 2, pages 195–214, 1991.

[39] R. Boyd and P. J. Richerson. Group Selection Among Alternative Evolutionarily Stable Strategies. *Theoretical Biology*, 145(3):331–342, 1990. ISSN: 0022-5193. doi: https://doi.org/10.1016/S0022-5193(05)80113-4.

[40] R. Boyd and P. J. Richerson. Culture and the Evolution of Human Cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3281–3288, 11 2009. ISSN: 0800-4622. doi: https://doi.org/10.1098/rstb.2009.0134.

[41] R. Boyd, P. J. Richerson, and J. Henrich. The Cultural Niche: Why Social Learning is Essential for Human Adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 2(Supplement 2):10918–25, 6 2011. doi: https://doi.org/10.1073/pnas.1100290108.

[42] V. A. Braithwaite and A. G. V. Salvanes. Environmental Variability in the Early Rearing Environment Generates Behaviourally Flexible Cod: Implications for Rehabilitating Wild Populations. *Proceedings of the Royal Society B: Biological Sciences*, 272(1568):1107–1113, 2005. doi: https://doi.org/10.1098/rspb.2005.3062.

[43] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo. Swarm Robotics: A Review from the Swarm Engineering Perspective. *Swarm Intelligence*, 7(1):1–41, 3 2013. doi: https://doi.org/10.1007/s11721-012-0075-2.

[44] L. Buşoniu, R. Babuška, and B. De Schutter. A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172, 2008. doi: https://doi.org/10.1109/TSMCC.2007.913919.

[45] J. A. Bullinaria. Understanding the Emergence of Modularity in Neural Systems. *Cognitive Science*, 2007. ISSN: 03640213. doi: https://doi.org/10.1080/15326900701399939.

[46] A. Burger, D. W. King, and G. Schiele. Reconfigurable Embedded Devices Using Reinforcement Learning to Develop Action-Policies. In *Proceedings of the 1st IEEE International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 232–241, 2020. doi: https://doi.org/10.1109/ACSOS49614.2020.00046.

[47] R. W. Byrne. Machiavellian Intelligence. *Evolutionary Anthropology: Issues, News, and Reviews*, 5(5):172–180, 1 1996. ISSN: 10601538. doi: https://doi.org/10.1002/(SICI)1520-6505(1996)5:5<172::AID-EVAN6>3.0.CO;2-H.

[48] G. Capi. Multiobjective Evolution of Neural Controllers and Task Complexity. *IEEE Transactions on Robotics*, 23(6):1225–1234, 2007. doi: https://doi.org/10.1109/TRO.2007.910773.

[49] A. Carter, A. Goldizen, and R. Heinsohn. Personality and Plasticity: Temporal Behavioural Reaction Norms in a Lizard, the Namibian Rock Agama. *Animal Behaviour*, 84(2):471–477, 2012. ISSN: 0003-3472. doi: https://doi.org/10.1016/j.anbehav.2012.06.001.

[50] J. T. Carvalho and S. Nolfi. Behavioural Plasticity in Evolving Robots. *Theory in Biosciences*, 135 (4):201–216, 2016. doi: https://doi.org/10.1007/s12064-016-0233-y.

[51] C. Castelfranchi. Modelling Social Action for AI Agents. *Artificial Intelligence*, 103(1-2):157–182, 8 1998. doi: https://doi.org/10.1016/S0004-3702(98)00056-3.

[52] C. Castelfranchi, F. Dignum, C. M. Jonker, and J. Treur. Deliberative Normative Agents: Principles and Architecture. In N. Jennings and Lespérance Y., editors, *Intelligent Agents VI. Agent Theories, Architecture, and Languages*, pages 364–378. Springer, Berlin, Heidelberg, 2000. doi: https://doi.org/10.1007/10719619_27.

[53] S. Chalotra, S. K. Sehra, and S. S. Sehra. A Systematic Review of Applications of Bee Colony Optimization. In *Proceedings of the International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, pages 257–260. IEEE, 2 2016. ISBN: 978-1-5090-2084-3. doi: https://doi.org/10.1109/ICICCS.2016.7542297.

[54] R. Cheng, M. J. Khojasteh, A. D. Ames, and J. W. Burdick. Safe Multi-Agent Interaction through Robust Control Barrier Functions with Learned Uncertainties. In *Proceedings of the 59th IEEE Conference on Decision and Control (CDC)*, pages 777–783, 2020. doi: https://doi.org/10.1109/CDC42340.2020.9304395.

[55] C. Cioffi-Revilla and M. Rouleau. MASON RebeLand: An Agent-Based Model of Politics, Environment, and Insurgency. *International Studies Review*, 12(1):31–52, 03 2010. ISSN: 1521-9488. doi: https://doi.org/10.1111/j.1468-2486.2009.00911.x.

[56] D. Cliff and L. Northrop. The Global Financial Markets: An Ultra-Large-Scale Systems Perspective. In *Large-Scale Complex IT Systems. Development, Operation and Management*, pages 29–70. Springer, 2012. doi: https://doi.org/10.1007/978-3-642-34059-8_2.

[57] H. G. Cobb and J. J. Grefenstette. Genetic Algorithms for Tracking Changing Environments. In *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 523–530, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc. ISBN: 1-55860-299-2, URL: http://dl.acm.org/citation.cfm?id=645513.657576.

[58] J. Cohen. *Statistical Power Analysis for the Behavioural Sciences*. Lawrence Erlbaum Associates, 2nd edition, 1988.

[59] A. Corl, K. Bi, C. Luke, A. S. Challa, A. J. Stern, B. Sinervo, and R. Nielsen. The Genetic Basis of Adaptation following Plastic Changes in Coloration in a Novel Environment. *Current Biology*, 28(18): 2970–2977, 2018. ISSN: 0960-9822. doi: https://doi.org/10.1016/j.cub.2018.06.075.

[60] A. T. Crooks and S. Wise. GIS and Agent-Based Models for Humanitarian Assistance. *Computers, Environment and Urban Systems*, 41:100–111, 2013. ISSN: 0198-9715. doi: https://doi.org/10.1016/j.compenvurbsys.2013.05.003.

[61] A. R. Daram, D. Kudithipudi, and A. Yanguas-Gil. Task-Based Neuromodulation Architecture for Lifelong Learning. In *Proceedings of the 20th International Symposium on Quality Electronic Design (ISQED)*, pages 191–197, 2019. doi: https://doi.org/10.1109/ISQED.2019.8697362.

[62] K. Dautenhahn. Ants Don't Have Friends – Thoughts on Socially Intelligent Agents. *Socially Intelligent Agents*, 97(2):22–27, 1997.

[63] K. Dautenhahn, B. Ogden, and T. Quick. From Embodied to Socially Embedded Agents — Implications for Interaction-Aware Robots. *Cognitive Systems Research*, 3(3):397–428, 2002. ISSN: 1389-0417. doi: https://doi.org/10.1016/S1389-0417(02)00050-5. Situated and Embodied Cognition.

[64] R. M. Dawes. Social Dilemmas. *Annual Review of Psychology*, 31(1):169–193, 1980.

[65] L. T. DeCarlo. On the Meaning and Use of Kurtosis. *Psychological Methods*, 2(3):292–307, 1997.

[66] A. Dezfouli and B. W. Balleine. Learning the Structure of the World: The Adaptive Nature of State-Space and Action Representations in Multi-Stage Decision-Making. *PLOS Computational Biology*, 15 (9):1–22, 09 2019. doi: https://doi.org/10.1371/journal.pcbi.1007334.

[67] E. Diener and T. K. Srull. Self-Awareness, Psychological Perspective, and Self-Reinforcement in Relation to Personal and Social Standards. *Personality and Social Psychology*, 37(3):413–423, 1979.

[68] F. Dignum, D. Morley, E. Sonenberg, and L. Cavedon. Towards Socially Sophisticated BDI Agents. In *Proceedings of the 4th International Conference on Multi Agent Systems*, pages 111–118. IEEE Comput. Soc, 2000. ISBN: 0-7695-0625-9. doi: https://doi.org/10.1109/ICMAS.2000.858442.

[69] N. J. Dingemanse and M. Wolf. Between-Individual Differences in Behavioural Plasticity Within Populations: Causes and Consequences. *Animal Behaviour*, 85(5):1031–1039, 2013. ISSN: 0003-3472. doi: https://doi.org/10.1016/j.anbehav.2012.12.032. Including Special Section: Behavioural Plasticity and Evolution.

[70] M. Doebeli and C. Hauert. Models of Cooperation Based on the Prisoner's Dilemma and the Snowdrift Game, 2005. ISSN: 1461023X.

[71] M. Dorigo, V. Maniezzo, and A. Colorni. Ant System: Optimization by a Colony of Cooperating Agents. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 26(1):29–41, 1996.

[72] R. Dreżewski. A Model of Co-evolution in Multi-agent System. In V. Mařík, M. Pěchouček, and J. Müller, editors, *Multi-Agent Systems and Applications III*, pages 314–323, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. ISBN: 978-3-540-45023-8. doi: https://doi.org/10.1007/3-540-45023-8_30.

[73] S. Ducatez, D. Sol, F. Sayol, and L. Lefebvre. Behavioural Plasticity is Associated with Reduced Extinction Risk in Birds. *Nature Ecology & Evolution*, 4(6):788–793, 2020. doi: https://doi.org/10.1038/s41559-020-1168-8.

[74] B. Duffy. Robots Social Embodiment in Autonomous Mobile Robotics. *International Journal of Advanced Robotic Systems*, 1(3):155–170, 2004. doi: https://doi.org/10.5772/5632.

[75] R. I. M. Dunbar. The Social Brain Hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5):178–190, 1 1998. ISSN: 10601538. doi: https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8.

[76] E. H. Durfee and T. A. Montgomery. MICE: A Flexible Testbed for Intelligent Coordination Experiments. In *Proceedings of the 1989 Distributed AI Workshop*, pages 25–40, 1989.

[77] G. Edwin and M. T. Cox. Resource Coordination in Single Agent and Multiagent Systems. In *Proceedings of the 13th IEEE International Conference on Tools with Artificial Intelligence (ICTAI) 2001*, pages 18–24, 2001. doi: https://doi.org/10.1109/ICTAI.2001.974444.

[78] K. O. Ellefsen, J. B. Mouret, and J. Clune. Neural Modularity Helps Organisms Evolve to Learn New Skills without Forgetting Old Skills. *PLoS Computational Biology*, 11(4):1–24, 04 2015. ISSN: 15537358. doi: https://doi.org/10.1371/journal.pcbi.1004128.

[79] J. I. Espinosa-Ramos, E. Capecci, and N. Kasabov. A Computational Model of Neuroreceptor-Dependent Plasticity (NRDP) Based on Spiking Neural Networks. *IEEE Transactions on Cognitive and Developmental Systems*, 11(1):63–72, 3 2019. ISSN: 2379-8939. doi: https://doi.org/10.1109/TCDS.2017.2776863.

[80] L. Esterle. Goal-Aware Team Affiliation in Collectives of Autonomous Robots. In *Proceedings of the IEEE 12th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 90–99. IEEE, 9 2018. ISBN: 978-1-5386-5172-8. doi: https://doi.org/10.1109/SASO.2018.00020.

[81] L. Esterle and J. N. Brown. Levels of Networked Self-Awareness. In *Proceedings of the IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 237–238. IEEE, 9 2018. doi: https://doi.org/10.1109/FAS-W.2018.00054.

[82] M. Fasli. From Social Agents to Multi-agent Systems: Preliminary Report. In *Proceedings of the International Central and Eastern European Conference on Multi-Agent Systems*, pages 111–121. Springer, 2003. ISBN: 978-3-540-45023-8.

[83] K. Fehl, D. J. van der Post, and D. Semmann. Co-evolution of Behaviour and Social Network Structure Promotes uman Cooperation. *Ecology Letters*, 14(6):546–551, 2011. doi: https://doi.org/10.1111/j.1461-0248.2011.01615.x.

[84] X. Fei, H. S. Mahmassani, and P. Murray-Tuite. Vehicular Network Sensor Placement Optimization Under Uncertainty. *Transportation Research Part C*, 29:14–31, 2013. doi: https://doi.org/10.1016/j.trc.2013.01.004.

[85] S. G. Ficici and J. B. Pollack. A Game-Theoretic Approach to the Simple Coevolutionary Algorithm. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature PPSN VI*, pages 467–476, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg. ISBN: 978-3-540-45356-7". doi: https://doi.org/10.1007/3-540-45356-3_46.

[86] S. G. Ficici and J. B. Pollack. Effects of Finite Populations on Evolutionary Stable Strategies. In *Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation*, pages 927–934, 2000.

[87] A. Field. *Discovering Statistics Using IBM SPSS Statistics*. SAGE, 2013.

[88] M. E. Ford and M. S. Tisak. A Further Search for Social Intelligence. *Educational Psychology*, 75(2): 196–206, 1983.

[89] S. I. F. Forss, E. Willems, J. Call, and C. P. van Schaik. Cognitive Differences between Orang-utan Species: A Test of the Cultural Intelligence Hypothesis. *Scientific Reports*, 6(1):30516, 9 2016. doi: https://doi.org/10.1038/srep30516.

[90] S. A. Foster. Evolution of Behavioural Phenotypes: Influences of Ancestry and Expression. *Animal Behaviour*, 85(5):1061–1075, 2013. ISSN: 0003-3472. doi: https://doi.org/10.1016/j.anbehav.2013.02.008. Including Special Section: Behavioural Plasticity and Evolution.

[91] U. Frith and C. Frith. The Social Brain: Allowing Humans to Boldly Go Where No Other Species Has Been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):165–176, 1 2010. ISSN: 0962-8436. doi: https://doi.org/10.1098/rstb.2009.0160.

[92] H. Fu, P. R. Lewis, B. Sendhoff, K. Tang, and X. Yao. What Are Dynamic Optimization Problems? In *Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC)*, pages 1550–1557, 2014. doi: https://doi.org/10.1109/CEC.2014.6900316.

[93] X. Fu and X. Gao. Genetic Algorithm with Adaptive Immigrants for Dynamic Flight Path Planning. In *Proceedings of the 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, volume 1, pages 630–634. IEEE, 2010. doi: https://doi.org/10.1109/ICICISYS.2010.5658510.

[94] G. G. Gallup. Self-Awareness and the Evolution of Social Intelligence. *Behavioural Processes*, 1998. ISSN: 03766357. doi: https://doi.org/10.1016/S0376-6357(97)00079-X.

[95] S. Garnier. *viridis: Default Color Maps from 'matplotlib'*, 2018, URL: https://CRAN.R-project.org/package=viridis. R package version 0.5.1.

[96] A. Ghouri, C. M. Barnes, and P. R. Lewis. A Minimal River Crossing Task to Aid the Explainability of Evolutionary Agents. *Artificial Life Conference Proceedings*, pages 36–43, 2020. doi: https://doi.org/10.1162/isal_a_00347.

[97] J. E. Godoy, I. Karamouzas, S. J. Guy, and M. Gini. Implicit Coordination in Crowded Multi-Agent Navigation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016. doi: https://doi.org/10.5555/3016100.3016247.

[98] F. Gomez and R. Miikkulainen. Incremental Evolution Of Complex General Behavior. *Adaptive Behavior*, 5(3–4):317–342, 1997, URL: http://nn.cs.utexas.edu/?gomez:ab97.

[99] W. S. Grant, J. Tanner, and L. Itti. Biologically Plausible Learning in Neural Networks with Modulatory Feedback. *Neural Networks*, 2017. ISSN: 18792782. doi: https://doi.org/10.1016/j.neunet.2017.01.007.

[100] S. Grossberg. Nonlinear Neural Networks: Principles, Mechanisms, and Architectures. *Neural Networks*, 1:17–61, 1988. ISSN: 08936080. doi: https://doi.org/10.1016/0893-6080(88)90021-4.

[101] J. Hähner, U. Brinkschulte, P. Lukowicz, S. Mostaghim, B. Sick, and S. Tomforde. Runtime Self-Integration as Key Challenge for Mastering Interwoven Systems. In *Proceedings of the 28th International Conference on Architecture of Computing Systems (ARCS)*, pages 1–8. VDE, 2015.

[102] A. W. Hamood and E. Marder. Animal-to-Animal Variability in Neuromodulation and Circuit Function. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 79, pages 21–28. Cold Spring Harbor Laboratory Press, 2014. doi: https://doi.org/10.1101/sqb.2014.79.024828.

[103] G. Herczeg, T. J. Urszán, S. Orf, G. Nagy, A. Kotrschal, and N. Kolm. Brain Size Predicts Behavioural Plasticity in Guppies (*Poecilia reticulata*): An Experiment. *Journal of Evolutionary Biology*, 32(3):218–226, 2019. doi: https://doi.org/10.5061/dryad.fp11572.

[104] C. Herpson. *Dedale: Demonstrating a Realistic Testbed for Decentralized Multi-agents Problems*, pages 426–429. Springer International Publishing, Cham, 2020. ISBN: 978-3-030-49778-1. doi: https://doi.org/10.1007/978-3-030-49778-1_40.

[105] E. Herrmann, J. Call, M. V. Hernàndez-Lloreda, B. Hare, and M. Tomasello. Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science*, 317(5843):1360–1366, 9 2007. ISSN: 1095-9203. doi: https://doi.org/10.1126/science.1146282.

[106] T. Hesselberg. The Mechanism Behind Plasticity of Web-Building Behavior in an Orb Spider Facing Spatial Constraints. *Journal of Arachnology*, 42(3):311–314, 2014. ISSN: 01618202, 19372396. doi: https://doi.org/10.1636/J14-05.1.

[107] F. Heylighen. Self-organization in Communicating Groups: The Emergence of Coordination, Shared References and Collective Intelligence. In À. Massip-Bonet and A. Bastardas-Boada, editors, *Complexity Perspectives on Language, Communication and Society*, pages 117–149. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN: 978-3-642-32817-6. doi: https://doi.org/10.1007/978-3-642-32817-6_10.

[108] B. Hrolenok, S. Luke, K. Sullivan, and C. Vo. Collaborative Foraging using Beacons. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, volume 10, pages 1197–1204, 2010. doi: https://doi.org/10.5555/1838186.1838193.

[109] C. Y. Huang, S. W. Wang, and C. T. Sun. Self-Aware Intelligent Agents in the Prisoner's Dilemma. In *Proceedings of the 2011 International Conference on Future Computer Sciences and Application, ICFCSA 2011*, pages 127–131, 2011. ISBN: 9780769544229. doi: https://doi.org/10.1109/ICFCSA.2011.36.

[110] J. Huang, X. Ruan, N. Yu, Q. Fan, J. Li, and J. Cai. A Cognitive Model Based on Neuromodulated Plasticity. *Computational Intelligence and Neuroscience*, 2016. ISSN: 16875273. doi: https://doi.org/10.1155/2016/4296356.

[111] T. Jansen and C. Zarges. Evolutionary Algorithms and Artificial Immune Systems on a Bi-Stable Dynamic Optimisation Problem. In *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*, GECCO '14, pages 975–982, New York, NY, USA, 2014. Association for Computing Machinery. ISBN: 9781450326629. doi: https://doi.org/10.1145/2576768.2598344.

[112] N. Jennings. Controlling Cooperative Problem Solving in Industrial Multi-Agent Systems using Joint Intentions. *Artificial Intelligence*, 75(2):195–240, 6 1995. ISSN: 0004-3702. doi: https://doi.org/10.1016/0004-3702(94)00020-2.

[113] Y. Jin and J. Branke. Evolutionary Optimization in Uncertain Environments – A Survey. *IEEE Transactions on Evolutionary Computation*, 9(3):303–317, 2005. doi: https://doi.org/10.1109/TEVC.2005.846356.

[114] B. Jolley and A. Channon. Toward Evolving Robust, Deliberate Motion Planning with HyperNEAT. In *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8, 2017. doi: https://doi.org/10.1109/SSCI.2017.8285284.

[115] B. Jolley and A. Channon. Evolving Robust, Deliberate Motion Planning With a Shallow Convolutional Neural Network. In *Artificial Life Conference Proceedings*, pages 536–543. MIT Press, 2018. doi: https://doi.org/10.1162/isal_a_00099.

[116] B. P. Jolley, J. M. Borg, and A. Channon. Analysis of Social Learning Strategies when Discovering and Maintaining Behaviours Inaccessible to Incremental Genetic Evolution. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9825 LNCS, pages 293–304, 2016. ISBN: 9783319434872. doi: https://doi.org/10.1007/978-3-319-43488-9_26.

[117] A. Jolly. Lemur Social Behavior and Primate Intelligence. *Science*, 153(3735):501–506, 7 1966. ISSN: 0036-8075. doi: https://doi.org/10.1126/science.153.3735.501.

[118] J. F. Kamhi, W. Gronenberg, S. K. A. Robson, and J. F. A. Traniello. Social Complexity Influences Brain Investment and Neural Operation Costs in Ants. *Proceedings of the Royal Society B: Biological Sciences*, 283(1841):20161949, 10 2016. doi: https://doi.org/10.1098/rspb.2016.1949.

[119] D. Karaboga and B. Basturk. A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm. *Journal of Global Optimization*, 39(3):459–471, 2007. doi: https://doi.org/10.1007/s10898-007-9149-x.

[120] Z. P. Kilpatrick, W. R. Holmes, T. L. Eissa, and K. Josić. Optimal Models of Decision-Making in Dynamic Environments. *Current Opinion in Neurobiology*, 58:54–60, 2019. ISSN: 0959-4388. doi: https://doi.org/10.1016/j.conb.2019.06.006. Computational Neuroscience.

[121] P. Kollock. Social Dilemmas: The Anatomy of Cooperation. *Annual Review of Sociology*, 1998. ISSN: 0360-0572. doi: https://doi.org/10.1146/annurev.soc.24.1.183.

[122] P. E. Komers. Behavioural Plasticity in Variable Environments. *Canadian Journal of Zoology*, 75(2): 161–169, 1997.

[123] L. Komsta and F. Novomestky. *moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests*, 2015, URL: https://CRAN.R-project.org/package=moments. R package version 0.14.

[124] M. Koren and S. Tenreyro. Volatility and Development. *The Quarterly Journal of Economics*, 122(1): 243–287, 2007. doi: https://doi.org/10.1162/qjec.122.1.243.

[125] S. Kounev, P. Lewis, K. L. Bellman, N. Bencomo, J. Camara, A. Diaconescu, L. Esterle, K. Geihs, H. Giese, S. Götz, P. Inverardi, J. O. Kephart, and A. Zisman. The Notion of Self-aware Computing. In *Self-Aware Computing Systems*, pages 3–16. Springer, 2017. ISBN: 978-3-319-47474-8. doi: https://doi.org/10.1007/978-3-319-47474-8_1.

[126] K. Kravari and N. Bassiliades. A Survey of Agent Platforms. *Journal of Artificial Societies and Social Simulation*, 18(1):11, 2015. ISSN: 1460–7425. doi: https://doi.org/10.18564/jasss.2661.

[127] J. L. Krichmar. The Neuromodulatory System: A Framework for Survival and Adaptive Behavior in a Challenging World. *Adaptive Behavior*, 16(6):385–399, 2008. doi: https://doi.org/10.1177/1059712308095775.

[128] Y. Kubera, P. Mathieu, and S. Picault. IODA: An Interaction-Oriented Approach for Multi-Agent Based Simulations. *Autonomous Agents and Multi-Agent Systems*, 23(3):303–343, 2011. doi: https://doi.org/10.1007/s10458-010-9164-z.

[129] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. Handwritten Digit Recognition with a Back-Propagation Network. In *Advances in Neural Information Processing Systems*, pages 396–404, 1990.

[130] S.-K. Lee and B.-R. Moon. Genetic Algorithm with Adaptive Elitism-Based Immigrants for Dynamic Optimization Problems. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, pages 1865–1866, 2009.

[131] J. Z. Leibo, V. Zambaldi, M. Lanctot, J. Marecki, and T. Graepel. Multi-agent Reinforcement Learning in Sequential Social Dilemmas. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473, 2017. doi: https://doi.org/10.5555/3091125.3091194.

[132] P. R. Lewis. Self-Aware Computing Systems: From Psychology to Engineering. In *Proceedings of the 2017 Design, Automation and Test in Europe, DATE 2017*, pages 1044–1049, 2017. ISBN: 9783981537093. doi: https://doi.org/10.23919/DATE.2017.7927144.

[133] P. R. Lewis, A. Chandra, and K. Glette. Self-awareness and Self-expression: Inspiration from Psychology. In P. R. Lewis, M. Platzner, B. Rinner, J. Torresen, and X. Yao, editors, *Self-aware Computing Systems*, chapter 2, pages 9–21. Springer International Publishing, 1 edition, 2016. ISBN: 978-3-319-39674-3. doi: https://doi.org/10.1007/978-3-319-39675-0.

[134] J. Li, G. Kendall, and R. John. Computing Nash Equilibria and Evolutionarily Stable States of Evolutionary Games. *IEEE Transactions on Evolutionary Computation*, 20(3):460–469, 2016. doi: https://doi.org/10.1109/TEVC.2015.2490076.

[135] M. Lihoreau, T. Latty, and L. Chittka. An Exploration of the Social Brain Hypothesis in Insects. *Frontiers in Physiology*, 3:442, 11 2012. ISSN: 1664-042X. doi: https://doi.org/10.3389/fphys.2012.00442.

[136] J. Lindblom and T. Ziemke. Social Situatedness: Vygotsky and Beyond. *Proceedings of the 2nd International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pages 71–78, 2002.

[137] S. Luke, G. C. Balan, L. Panait, C. Cioffi-Revilla, and S. Paus. MASON: A Java Multi-Agent Simulation Library. In *Proceedings of Agent 2003 Conference on Challenges in Social Simulation*, 2003.

[138] S. Luke, C. Cioffi-Revilla, L. Panait, and K. Sullivan. MASON: A New Multi-Agent Simulation Toolkit. In *Proceedings of the 2004 SwarmFest Workshop*, pages 316–327. Michigan, USA, 2004.

[139] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. MASON: A Multi-Agent Simulation Environment. *Simulation*, 81(7):517–527, 2005.

[140] A. Mahdavi and M. Carvalho. Distributed Coordination of Autonomous Guided Vehicles in Multi-agent Systems with Shared Resources. In *2019 SoutheastCon*, pages 1–7, 2019. doi: https://doi.org/10.1109/SoutheastCon42311.2019.9020456.

[141] S. Malakuti. Detecting Emergent Interference in Integration of Multiple Self-Adaptive Systems. In *Proceedings of the 2014 European Conference on Software Architecture Workshops*, ECSAW '14, New York, NY, USA, 2014. Association for Computing Machinery. ISBN: 9781450327787. doi: https://doi.org/10.1145/2642803.2642826.

[142] S. Mangiafico. *rcompanion: Functions to Support Extension Education Program Evaluation*, 2021, URL: https://CRAN.R-project.org/package=rcompanion. R package version 2.3.27.

[143] N. Y. Masse, G. D. Grant, and D. J. Freedman. Alleviating Catastrophic Forgetting using Context-Dependent Gating and Synaptic Stabilization. *Proceedings of the National Academy of Sciences*, 115 (44):E10467–E10475, 2018. ISSN: 0027-8424. doi: https://doi.org/10.1073/pnas.1803839115.

[144] M. J. Matarić. *Interaction and Intelligent Behavior*. PhD thesis, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science, 1994, URL: https://dspace.mit.edu/bitstream/handle/1721.1/7343/AITR-1495.pdf?sequence=2&isAllowed=y.

[145] P. McBurney. What are Models for? In M. Cossentino, M. Kaisers, K. Tuyls, and G. Weiß, editors, *Multi-Agent Systems. EUMAS 2011*, volume 7541 of *Lecture Notes in Computer Science*, pages 175–188, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN: 978-3-642-34799-3.

[146] M. McCloskey and N. J. Cohen. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. *Psychology of Learning and Motivation - Advances in Research and Theory*, 1989. ISSN: 00797421. doi: https://doi.org/10.1016/S0079-7421(08)60536-8.

[147] R. K. Merton. The Unanticipated Consequences of Purposive Social Action. *American Sociological Review*, 1936. ISSN: 00031224. doi: https://doi.org/10.2307/2084615.

[148] F. Mery and J. G. Burns. Behavioural Plasticity: An Interaction between Evolution and Experience. *Evolutionary Ecology*, 24(3):571–583, 2010. ISSN: 02697653. doi: https://doi.org/10.1007/s10682-009-9336-y.

[149] A. Mesoudi, L. Chang, S. R. Dall, and A. Thornton. The Evolution of Individual and Cultural Variation in Social Learning. *Trends in Ecology & Evolution*, 31(3):215–225, 3 2016. ISSN: 0169-5347. doi: https://doi.org/10.1016/J.TREE.2015.12.012.

[150] N. Michael, J. Fink, and V. Kumar. Experimental Testbed for Large Multirobot Teams. *IEEE Robotics Automation Magazine*, 15(1):53–61, 2008. doi: https://doi.org/10.1109/M-RA.2007.914924.

[151] T. Miconi. Evosphere: Evolutionary Dynamics in a Population of Fighting Virtual Creatures. In *Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 3066–3073, 2008. doi: https://doi.org/10.1109/CEC.2008.4631212.

[152] K. Mogielski and T. Płatkowski. A Mechanism of Dynamical Interactions for Two-Person Social Dilemmas. *Journal of Theoretical Biology*, 2009. ISSN: 00225193. doi: https://doi.org/10.1016/j.jtbi.2009.06.007.

[153] H. Moll and M. Tomasello. Cooperation and Human Cognition: The Vygotskian Intelligence Hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):639–648, 4 2007. ISSN: 0962-8436. doi: https://doi.org/10.1098/rstb.2006.2000.

[154] F. Mondada, L. Maria Gambardella, D. Floreano, S. Nolfi, J.-L. Deneubourg, and M. Dorigo. The Cooperation of Swarm-Bots: Physical Interactions in Collective Robotics. *IEEE Robotics Automation Magazine*, 12(2):21–28, 6 2005. doi: https://doi.org/10.1109/MRA.2005.1458313.

[155] C. Müller-Schloer. Organic Computing – On the Feasibility of Controlled Emergence. *Proceedings of the 2nd IEEE/ACM/IFIP international conference on Hardware/Software Codesign and System Synthesis*, 2004. doi: https://doi.org/10.1145/1016720.1016724.

[156] M. Muthukrishna, M. Doebeli, M. Chudek, and J. Henrich. The Cultural Brain Hypothesis: How Culture Drives Brain Expansion, Sociality, and Life History. *PLoS computational biology*, 14(11): e1006504, 2018. doi: https://doi.org/10.1371/journal.pcbi.1006504.

[157] V. Nallur and S. Clarke. Clonal Plasticity: An Autonomic Mechanism for Multi-Agent Systems to Self-Diversify. *Autonomous Agents and Multi-Agent Systems*, 32(2):275–311, 2018. doi: https://doi.org/10.1007/s10458-017-9380-x.

[158] K. K. Ndousse, D. Eck, S. Levine, and N. Jaques. Emergent Social Learning via Multi-agent Reinforcement Learning. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7991–8004. PMLR, 7 2021, URL: http://proceedings.mlr.press/v139/ndousse21a.html. Pre-print available at: https://arxiv.org/pdf/2010.00581v2.pdf.

[159] P. R. Nelson. Ensuring Appropriate Use of Computational Resources in Interwoven Systems. In *Proceedings of the IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, pages 139–143. IEEE, 2017. doi: https://doi.org/10.1109/FAS-W.2017.135.

[160] D. Ngar-yin Mah, J. M. van der Vleuten, J. Chi-man Ip, and P. Ronald Hills. Governing the Transition of Socio-technical Systems: A Case Study of the Development of Smart Grids in Korea. *Energy Policy*, 45:133–141, 6 2012. doi: https://doi.org/10.1016/j.enpol.2012.02.005.

[161] T. T. Nguyen, S. Yang, and J. Branke. Evolutionary Dynamic Optimization: A Survey of the State of the Art. *Swarm and Evolutionary Computation*, 6:1–24, 2012. ISSN: 2210-6502. doi: https://doi.org/10.1016/j.swevo.2012.05.001.

[162] M. Nielsen. Copying Actions and Copying Outcomes: Social Learning through the Second Year. *Developmental Psychology*, 2006. ISSN: 00121649. doi: https://doi.org/10.1037/0012-1649.42.3.555.

[163] R. F. Oliveira. Social Plasticity in Fish: Integrating Mechanisms and Function. *Journal of Fish Biology*, 81(7):2127–2150, 2012. doi: https://doi.org/10.1111/j.1095-8649.2012.03477.x.

[164] S. Ossowski and A. García-Serrano. Social Structure in Artificial Agent Societies: Implications for Autonomous Problem-Solving Agents. In *Proceedings of the 5th International Workshop on Agent Theories, Architectures and Languages*, pages 133–148, 1998. ISBN: 978-3-540-65713-2. doi: https://doi.org/10.1007/3-540-49057-4_9.

[165] L. Panait, R. P. Wiegand, and S. Luke. Improving Coevolutionary Search for Optimal Multiagent Behaviors. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, pages 653—-658, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc. doi: https://doi.org/10.5555/1630659.1630755.

[166] J. Pavón, M. Arroyo, S. Hassan, and C. Sansores. Agent-Based Modelling and Simulation for the Analysis of Social Patterns. *Pattern Recognition Letters*, 29(8):1039–1048, 2008. ISSN: 0167-8655. doi: https://doi.org/10.1016/j.patrec.2007.06.021. Pattern Recognition in Interdisciplinary Perception and Intelligence.

[167] C. Perreault, C. Moya, and R. Boyd. A Bayesian Approach to the Evolution of Social Learning. *Evolution and Human Behavior*, 33(5):449–459, 9 2012. ISSN: 1090-5138. doi: https://doi.org/10.1016/J.EVOLHUMBEHAV.2011.12.007.

[168] L. Pitonakova, R. Crowder, and S. Bullock. Information Flow Principles for Plasticity in Foraging Robot Swarms. *Swarm Intelligence*, 10(1):33–63, 2016. doi: https://doi.org/10.1007/s11721-016-0118-1.

[169] J. Pitt, J. Schaumeier, and A. Artikis. The Axiomatisation of Socio-Economic Principles for Self-Organising Systems. In *Proceedings of the 5th IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, pages 138–147. IEEE, 10 2011. doi: https://doi.org/10.1109/SASO.2011.25.

[170] S. H. Poon and C. W. Granger. Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, 41(2):478–539, 2003. doi: https://doi.org/10.1257/jel.41.2.478.

[171] S. T. Powers, A. Ekárt, and P. R. Lewis. Modelling Enduring Institutions: The Complementarity of Evolutionary and Agent-Based Approaches. *Cognitive Systems Research*, 52:67–81, 2018. ISSN: 1389-0417. doi: https://doi.org/10.1016/j.cogsys.2018.04.012.

[172] T. Qian, T. F. Jaeger, and R. Aslin. Learning to Represent a Multi-Context Environment: More than Detecting Changes. *Frontiers in Psychology*, 3:228, 2012. ISSN: 1664-1078. doi: https://doi.org/10.3389/fpsyg.2012.00228.

[173] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013, URL: http://www.R-project.org/.

[174] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation-based Negotiation. *The Knowledge Engineering Review*, 18(4):343–375, 2003.

[175] A. S. Rao, M. P. Georgeff, and E. A. Sonenberg. Social Plans: A Preliminary Report. In *Proceedings of the 3rd European Workshop on Modelling Autonomous Agents and Multi-Agent Worlds (MAAMAW-91)*, volume 3 of *Decentralized AI*, pages 57–76. Elsevier, 1992.

[176] S. M. Reader. Innovation and Social Learning: Individual Variation and Brain Evolution. *Animal Biology*, 2003. ISSN: 15707555. doi: https://doi.org/10.1163/157075603769700340.

[177] H. Ritz, M. R. Nassar, M. J. Frank, and A. Shenhav. A Control Theoretic Model of Adaptive Learning in Dynamic Environments. *Journal of Cognitive Neuroscience*, 30(10):1405–1421, 10 2018. ISSN: 0898-929X. doi: https://doi.org/10.1162/jocn_a_01289.

[178] E. Robinson, T. Ellis, and A. Channon. Neuroevolution of Agents Capable of Reactive and Deliberative Behaviours in Novel and Dynamic Environments. In *Advances in Artificial Life*, pages 1–10. Springer, 2007. ISBN: 9783540749127. doi: https://doi.org/10.1007/978-3-540-74913-4_35.

[179] S. G. Robinson, C. M. Barnes, and P. R. Lewis. Centralised and Decentralised Control of Video Game Agents. In *Proceedings of the 20th UK Workshop on Computational Intelligence*, 2021. In Press.

[180] E. Ronald and M. Schoenauer. Genetic Lander: An Experiment in Accurate Neuro-Genetic Control. In *Proceedings of the 3rd Conference on Parallel Problem Solving from Nature*, pages 452–461. Springer-Verlag, 1994.

[181] R. Rosenthal. *Meta-Analytic Procedures for Social Research*, volume 6 of *Applied Social Research Methods*. SAGE, 1991. doi: https://doi.org/10.4135/9781412984997.

[182] C. D. Rosin and R. K. Belew. New Methods for Competitive Coevolution. *Evolutionary Computation*, 5(1):1–29, 1997. doi: https://doi.org/10.1162/evco.1997.5.1.1.

[183] S. Rudolph, S. Tomforde, B. Sick, and J. Hähner. A Mutual Influence Detection Algorithm for Systems with Local Performance Measurement. In *Proceedings of the 9th IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, pages 144–149. IEEE, 9 2015. ISBN: 978-1-4673-7535-1. doi: https://doi.org/10.1109/SASO.2015.23.

[184] S. Russell. Learning Agents for Uncertain Environments. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pages 101–103, 1998.

[185] T. L. Rymer, N. Pillay, and C. Schradin. Extinction or Survival? Behavioral Flexibility in Response to Environmental Change in the African Striped Mouse Rhabdomys. *Sustainability*, 5(1):163–186, 2013. doi: https://doi.org/10.3390/su5010163.

[186] P. Savaget, M. Geissdoerfer, A. Kharrazi, and S. Evans. The Theoretical Foundations of Sociotechnical Systems Change for Sustainability: A Systematic Literature Review. *Journal of Cleaner Production*, 206:878–892, 2019. ISSN: 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2018.09.208.

[187] M. Schneider-Fontán and M. Mataric. Territorial Multi-Robot Task Division. *IEEE Transactions on Robotics and Automation*, 14(5):815–822, 1998. ISSN: 1042296X. doi: https://doi.org/10.1109/70.720357.

[188] D. J. Shaw, K. Czekóová, J. Chromec, R. Mareček, and M. Brázdil. Copying You Copying Me: Interpersonal Motor Co-ordination Influences Automatic Imitation. *PLOS ONE*, 2013. doi: https://doi.org/10.1371/journal.pone.0084820.

[189] K. Sigmund and M. A. Nowak. Evolutionary Game Theory. *Current Biology*, 1999. ISSN: 09609822. doi: https://doi.org/10.1016/S0960-9822(99)80321-2.

[190] A. Simões and E. Costa. An Immune System-Based Genetic Algorithm to Deal with Dynamic Environments: Diversity and Memory. In *Artificial Neural Nets and Genetic Algorithms*, pages 168–174. Springer, 2003.

[191] H. A. Simon. How Complex are Complex Systems? In *Proceedings of the 1976 Biennial Meeting of the Philosophy of Science Association*, volume 2, pages 507–522, 1976.

[192] J. A. Simpson and D. T. Kenrick. *Evolutionary Social Psychology*. Lawrence Erlbaum Associates, 1997. ISBN: 0805824200.

[193] P. E. Smaldino, J. C. Schank, and R. McElreath. Increased Costs of Cooperation Help Cooperators in the Long Run. *The American Naturalist*, 181(4):451–463, 2013. doi: https://doi.org/10.1086/669615.

[194] J. M. Smith. The Theory of Games and the Evolution of Animal Conflicts. *Journal of Theoretical Biology*, 47(1):209–221, 1974.

[195] J. M. Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[196] E. C. Snell-Rood. An Overview of the Evolutionary Causes and Consequences of Behavioural Plasticity. *Animal Behaviour*, 2013. ISSN: 00033472. doi: https://doi.org/10.1016/j.anbehav.2012.12.031.

[197] A. Soltoggio, J. A. Bullinaria, C. Mattiussi, P. Dürr, and D. Floreano. Evolutionary Advantages of Neuromodulated Plasticity in Dynamic, Reward-Based Scenarios. In *Proceedings of the 11th International Conference on the Simulation and Synthesis of Living Systems, ALIFE 2008*, 2008. ISBN: 9780262750172.

[198] F. M. J. Sommerlandt, A. Brockmann, W. Rössler, and J. Spaethe. Immediate Early Genes in Social Insects: A Tool to Identify Brain Regions Involved in Complex Behaviors and Molecular Processes Underlying Neuroplasticity. *Cellular and Molecular Life Sciences*, 76(4):637–651, 2019. doi: https://doi.org/10.1007/S00018-018-2948-Z.

[199] J. A. Stamps. Individual Differences in Behavioural Plasticities. *Biological Reviews*, 2016. ISSN: 1469185X. doi: https://doi.org/10.1111/brv.12186.

[200] K. O. Stanley and R. Miikkulainen. Evolving Neural Networks through Augmenting Topologies. *Evolutionary Computation*, 10(2):99–127, 2002. doi: https://doi.org/10.1162/106365602320169811.

[201] K. O. Stanley, D. B. D'Ambrosio, and J. Gauci. A Hypercube-Based Encoding for Evolving Large-Scale Neural Networks. *Artificial Life*, 15(2):185–212, 2009. doi: https://doi.org/10.1162/artl.2009.15.2.15202.

[202] K. O. Stanley, J. Clune, J. Lehman, and R. Miikkulainen. Designing Neural Networks through Neuroevolution. *Nature Machine Intelligence*, 1:25–35, 2019. doi: https://doi.org/10.1038/s42256-018-0006-z.

[203] A. Stanton and A. Channon. Incremental Neuroevolution of Reactive and Deliberative 3D Agents. *European Conference on Artificial Life (ECAL)*, pages 341–348, 2015. doi: https://doi.org/10.7551/978-0-262-33027-5-ch063.

[204] A. Stanton and A. Channon. Neuroevolution of Feedback Control for Object Manipulation by 3D Agents. In *Proceedings of the 15th International Conference on the Synthesis and Simulation of Living Systems*, pages 144–151, 07 2016. doi: https://doi.org/10.1162/978-0-262-33936-0-ch030.

[205] A. Stein, S. Tomforde, D. Rauh, and J. Hähner. Dealing with Unforeseen Situations in the Context of Self-Adaptive Urban Traffic Control: How to Bridge the Gap. In *Proceedings of the 2016 IEEE International Conference on Autonomic Computing (ICAC)*, pages 167–172, 2016. doi: https://doi.org/10.1109/ICAC.2016.20.

[206] S. E. Street, A. F. Navarrete, S. M. Reader, and K. N. Laland. Coevolution of Cultural Intelligence, Extended Life History, Sociality, and Brain Size in Primates. *Proceedings of the National Academy of Sciences of the United States of America*, 114(30):7908–7914, 7 2017.

[207] K. P. Sycara. Argumentation: Planning Other Agents' Plans. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence – Volume 1*, IJCAI'89, pages 517–523, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[208] G. Syswerda. Uniform Crossover in Genetic Algorithms. In *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 2–9, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc. ISBN: 1558600663. doi: https://doi.org/10.5555/645512.657265.

[209] P. Sztompka. *Socjologia*. Znak, Kraków, 2002.

[210] E. Taysom and N. Crilly. Resilience in Sociotechnical Systems: The Perspectives of Multiple Stakeholders. *She Ji: The Journal of Design, Economics, and Innovation*, 3(3):165–182, 2017. ISSN: 2405-8726. doi: https://doi.org/10.1016/j.sheji.2017.10.011.

[211] J. Thangarajah, L. Padgham, and M. Winikoff. Detecting & Avoiding Interference between Goals in Intelligent Agents. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 2003, URL: https://researchbank.rmit.edu.au/view/rmit%3A2589/n2003000372.pdf.

[212] J. Thøgersen. Social Norms and Cooperation in Real-Life Social Dilemmas. *Journal of Economic Psychology*, 2008. ISSN: 01674870. doi: https://doi.org/10.1016/j.joep.2007.12.004.

[213] S. Tisue and U. Wilensky. NetLogo: Design and Implementation of a Multi-Agent Modeling Environment. In *Proceedings of the Agent 2004 Conference*, 2004.

[214] M. Tomczak and E. Tomczak. The Need to Report Effect Size Estimates Revisited. An Overview of Some Recommended Measures of Effect Size. *Trends in Sport Sciences*, 21(1), 2014.

[215] S. Tomforde, S. Rudolph, K. Bellman, and R. Würtz. An Organic Computing Perspective on Self-Improving System Interweaving at Runtime. In *Proceedings of the 2016 IEEE International Conference on Autonomic Computing (ICAC)*, pages 276–284. IEEE, 2016. doi: https://doi.org/10.1109/ICAC.2016.15.

[216] K. Tsumori and S. Ozawa. Incremental Learning in Dynamic Environments Using Neural Network with Long-term Memory. In *Proceedings of the International Joint Conference on Neural Networks, 2003*, volume 4, pages 2583–2588. IEEE, 2003.

[217] U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission. Findings Regarding the Market Events of May 6, 2010, 2010. Available at: https://www.sec.gov/news/studies/2010/marketevents-report.pdf.

[218] P. A. van Lange, J. Joireman, C. D. Parks, and E. Van Dijk. The Psychology of Social Dilemmas: A Review. *Organizational Behavior and Human Decision Processes*, 2013. ISSN: 07495978. doi: https://doi.org/10.1016/j.obhdp.2012.11.003.

[219] C. P. van Schaik and J. M. Burkart. Social Learning and Evolution: The Cultural Intelligence Hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1567):1008–1016, 4 2011. ISSN: 1471-2970. doi: https://doi.org/10.1098/rstb.2010.0304.

[220] N. Vecoven, D. Ernst, A. Wehenkel, and G. Drion. Introducing Neuromodulation in Deep Neural Networks to Learn Adaptive Behaviours. *PLOS ONE*, 15(1):1–13, 01 2020. doi: https://doi.org/10.1371/journal.pone.0227922.

[221] F. Vega-Redondo. Expectations, Drift, and Volatility in Evolutionary Games. *Games and Economic Behavior*, 11(2):391–412, 1995.

[222] R. Velez and J. Clune. Diffusion-Based Neuromodulation can Eliminate Catastrophic Forgetting in Simple Neural Networks. *PLoS ONE*, 2017. ISSN: 19326203. doi: https://doi.org/10.1371/journal.pone.0187736.

[223] M. Viney and A. Diaz. Phenotypic Plasticity in Nematodes. *Worm*, 1(2):98–106, 2012. doi: https://doi.org/10.4161/worm.21086.

[224] L. S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, 1978. Edited by Michael Cole, Vera John-Steiner, Sylvia Scribner and Ellen Souberman.

[225] G. H. Walker, N. A. Stanton, D. Jenkins, P. Salmon, M. Young, and A. Aujla. Sociotechnical Theory and NEC System Design. In *Engineering Psychology and Cognitive Ergonomics*, pages 619–628. Springer, 2007. doi: https://doi.org/10.1007/978-3-540-73331-7_68.

[226] J. X. Wang, E. Hughes, C. Fernando, W. M. Czarnecki, E. A. Duéñez Guzmán, and J. Z. Leibo. Evolving Intrinsic Motivations for Altruistic Behavior. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, pages 683–692, Richland, SC, 2019. International Foundation for Autonomous Agents and Multiagent Systems. ISBN: 978-1-4503-6309-9. doi: https://doi.org/10.5555/3306127.3331756.

[227] M. Weber. *Economy and Society: An Outline of Interpretive Sociology*. University of California Press, 1978. ISBN: 0520035003. Edited by Guenther Roth and Claus Wittich.

[228] H. Wickham. Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12):1–20, 2007, URL: http://www.jstatsoft.org/v21/i12/.

[229] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4, URL: https://ggplot2.tidyverse.org.

[230] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: https://doi.org/10.21105/joss.01686.

[231] G. Widmer and M. Kubat. Effective Learning in Dynamic Environments by Explicit Context Tracking. In *Proceedings of the European Conference on Machine Learning*, pages 227–243. Springer, 1993.

[232] F. Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6):80–83, 1945. ISSN: 00994987, URL: http://www.jstor.org/stable/3001968.

[233] U. Wilensky. *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999, URL: http://ccl.northwestern.edu/netlogo/.

[234] H. Wolf. Volatility: Definitions and Consequences. In J. Aizenman and B. Pinto, editors, *Managing Economic Volatility and Crises: A Practitioner's Guide*, pages 45–64. Cambridge University Press, 2005. ISBN: 9780511510755. doi: https://doi.org/10.1017/CBO9780511510755.004.

[235] L. Yaeger. Computational Genetics, Physiology, Metabolism, Neural Systems, Learning, Vision, and Behavior or Polyworld: Life in a New Context. In C. G. Langton, editor, *Proceedings of the Artificial Life III Conference*, pages 263—298. Addison-Wesley, 1994.

[236] S. Yang. Genetic Algorithms with Memory- and Elitism-Based Immigrants in Dynamic Environments. *Evolutionary Computation*, 16(3):385–416, 2008. doi: https://doi.org/10.1162/evco.2008.16.3.385.

[237] S. X. Yang and M. Meng. An Efficient Neural Network Method for Real-Time Motion Planning with Safety Consideration. *Robotics and Autonomous Systems*, 32:115–128, 2000. ISSN: 09218890. doi: https://doi.org/10.1016/S0921-8890(99)00113-X.

[238] S. X. Yang and M. Meng. An Efficient Neural Network Approach to Dynamic Robot Motion Planning. *Neural Networks*, 13(2):143–148, 2000. ISSN: 08936080. doi: https://doi.org/10.1016/S0893-6080(99)00103-3.

[239] Z. Yang, K. Tang, and X. Yao. Large Scale Evolutionary Optimization using Cooperative Coevolution. *Information Sciences*, 178(15):2985–2999, 2008. ISSN: 0020-0255. doi: https://doi.org/10.1016/j.ins.2008.02.017.

[240] X. Yao. Evolving Artificial Neural Networks. *Proceedings of the IEEE*, 87(9):1423–1447, 9 1999. ISSN: 1558-2256. doi: https://doi.org/10.1109/5.784219.

[241] B. W. Yap and C. H. Sim. Comparisons of Various Types of Normality Tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, 2011. doi: https://doi.org/10.1080/00949655.2010.520163.

[242] J. Yoder and L. Yaeger. Evaluating Topological Models of Neuromodulation in Polyworld. In *Proceedings of the 14th International Conference on the Synthesis and Simulation of Living Systems*, pages 916–923. MIT Press, 2014. doi: https://doi.org/10.7551/978-0-262-32621-6-ch149.

[243] C. H. Yong and R. Miikkulainen. Coevolution of Role-Based Cooperation in Multiagent Systems. *IEEE Transactions on Autonomous Mental Development*, 1(3):170–186, 2009. doi: https://doi.org/10.1109/TAMD.2009.2037732.

[244] X. Yu, K. Tang, and X. Yao. An Immigrants Scheme Based on Environmental Information for Genetic Algorithms in Changing Environments. In *Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 1141–1147. IEEE, 2008. doi: https://doi.org/10.1109/CEC.2008.4630940.